

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 9, 2001, 11:45:45 ; Search time 16.02 seconds  
(without alignments)  
1521.783 Million cell updates/sec

Title: US-09-609-383-2

Perfect score: 6317  
Sequence: 1 MWGTRAWVSFLVLEVTSL.....QSGVSLRPRVAAQPLIN 1184

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Archived: 197339 segs, 20590346 residues

Actual number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA: \*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep: \*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep: \*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep: \*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep: \*  
5: /cgn2\_6/ptodata/2/1aa/PTUS.COMB.pep: \*  
6: /cgn2\_6/ptodata/2/1aa/backfile1.pep: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6297	99.7	1184	2 US-08-918-914-1	Sequence 1, Appli
2	6297	99.7	1184	3 US-08-996-083-3	Sequence 3, Appli
3	3091.5	48.9	1156	3 US-08-996-083-1	Sequence 1, Appli
4	3091.5	48.9	1156	4 US-09-429-516-1	Sequence 1, Appli
5	3091.5	48.9	1156	4 US-09-429-516-3	Sequence 3, Appli
6	2898.5	45.9	599	2 US-08-954-333-7	Sequence 7, Appli
7	157	2.5	1170	1 US-08-313-288B-20	Sequence 20, Appli
8	147.5	2.3	788	2 US-08-918-914-4	Sequence 4, Appli
9	144	2.3	1911	1 US-08-348-006B-5	Sequence 5, Appli
10	144	2.3	1911	2 US-08-800-825A-5	Sequence 5, Appli
11	144	2.3	1911	4 US-09-158-657-5	Sequence 5, Appli
12	144	2.3	1911	5 PCT-US94-10166-5	Sequence 5, Appli
13	143	2.3	1172	1 US-08-313-288B-19	Sequence 19, Appli
14	142.5	2.3	4544	1 US-08-469-486-52	Sequence 52, Appli
15	142.5	2.3	4544	2 US-08-469-486-52	Sequence 52, Appli
16	138	2.2	1101	2 US-08-469-537A-96	Sequence 96, Appli
17	136	2.2	1118	1 US-07-934-393B-2	Sequence 2, Appli
18	136	2.2	1118	1 US-08-278-089A-2	Sequence 2, Appli
19	136	2.2	1118	2 US-08-838-957A-2	Sequence 2, Appli
20	135.5	2.1	898	2 US-08-808-982-5	Sequence 5, Appli
21	135	2.1	1122	2 US-08-278-089A-6	Sequence 6, Appli
22	135	2.1	1122	1 US-08-838-957A-6	Sequence 6, Appli
23	132	2.1	218	3 US-08-985-526-1	Sequence 1, Appli
24	132	2.1	239	5 PCT-US93-01652-1	Sequence 1, Appli
25	130.5	2.1	644	5 PCT-US93-00031-21	Sequence 21, Appli
26	130.5	2.1	1101	3 US-08-986-485-2	Sequence 2, Appli
27	128.5	2.0	807	1 US-07-862-021B-10	Sequence 10, Appli

28	128.5	2.0	807	1 US-08-313-288B-10	Sequence 10, Appli
29	128.5	2.0	807	5 PCT-US93-03164-10	Sequence 10, Appli
30	128.5	2.0	1338	3 US-08-750-141A-3	Sequence 3, Appli
31	128	2.0	1501	2 US-08-447-464-3	Sequence 3, Appli
32	128	2.0	1501	2 US-08-716-679-3	Sequence 3, Appli
33	127	2.0	441	3 US-08-985-526-3	Sequence 3, Appli
34	126	2.0	736	5 PCT-US93-00031-15	Sequence 15, Appli
35	125	2.0	643	5 PCT-US93-00031-19	Sequence 19, Appli
36	124.5	2.0	1018	1 US-08-408-093-6	Sequence 6, Appli
37	124.5	2.0	1018	1 US-08-408-420A-6	Sequence 6, Appli
38	124.5	2.0	1018	1 US-08-714-901-6	Sequence 6, Appli
39	124.5	2.0	1018	3 US-08-040-741-6	Sequence 6, Appli
40	121	1.9	596	2 US-08-752-307B-13	Sequence 13, Appli
41	121	1.9	802	1 US-07-862-021B-12	Sequence 12, Appli
42	121	1.9	802	1 US-08-313-288B-12	Sequence 12, Appli
43	121	1.9	802	5 PCT-US93-03164-12	Sequence 12, Appli
44	120.5	1.9	735	5 PCT-US93-00031-13	Sequence 13, Appli
45	120	1.9	392	2 US-08-799-173A-7	Sequence 7, Appli

## ALIGNMENTS

RESULT 1  
US-08-918-914-1  
Sequence 1, Application US/08918914  
Patent No. 5876963  
GENERAL INFORMATION:  
APPLICANT: Mitchell, Peter  
APPLICANT: Hutchinson, Nancy  
APPLICANT: Lawton, Michael  
APPLICANT: Magna, Holly  
APPLICANT: Vocum, Sue  
APPLICANT: Murry, Lynn E.  
TITLE OF INVENTION: HUMAN NUCLEOTIDE PYROPHOSPHORYLASE  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESS: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Dr.  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/918,914  
FILING DATE: Filed Herewith  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0369  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
TELEX:  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1184 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: ???  
CLONE: 422069  
US-08-918-914-1

Query Match 99.7%; Score 6297; DB 2; Length 1184;  
 Best Local Similarity 99.7%; Pred. No. 0;  
 Matches 1181; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MGTGKAVFSEFLVLEVTSLVGRQTMLTQSVRRVOPGKKNSIFAKPADTLESFGEWTTLF 60  
 1 MGTGKAVFSEFLVLEVTSLVGRQTMLTQSVRRVOPGKKNSIFAKPADTLESFGEWTTLF 60  
 Db 1 MGTGKAVFSEFLVLEVTSLVGRQTMLTQSVRRVOPGKKNSIFAKPADTLESFGEWTTLF 60

QY 61 NIDYGGKGYERLDAIRFYGGDVCARPLRLERATDWTDPAGSTGOVHGSFREGFWCL 120  
 61 NIDYGGKGYERLDAIRFYGGDVCARPLRLERATDWTDPAGSTGOVHGSFREGFWCL 120  
 Db 61 NIDYGGKGYERLDAIRFYGGDVCARPLRLERATDWTDPAGSTGOVHGSFREGFWCL 120

QY 121 NREORPQNSNTVTRFLCPGSLRRDTERIWSFWSKSCSAACGQTVQTRTRICLAE 180  
 121 NREORPQNSNTVTRFLCPGSLRRDTERIWSFWSKSCSAACGQTVQTRTRICLAE 180  
 Db 121 NREORPQNSNTVTRFLCPGSLRRDTERIWSFWSKSCSAACGQTVQTRTRICLAE 180

QY 181 MVSICSEASEGQHKMGQDCTACDLTCPMGOVNADCDACMCOFMLHGAVSLPGAPASG 240  
 181 MVSICSEASEGQHKMGQDCTACDLTCPMGOVNADCDACMCOFMLHGAVSLPGAPASG 240  
 Db 181 MVSICSEASEGQHKMGQDCTACDLTCPMGOVNADCDACMCOFMLHGAVSLPGAPASG 240

QY 241 AAIYLLTKTKLTLQTDSDGRFRIPGLCPDGKSLIKTTKVFAPIVLTMTKTSLKAAATIK 300  
 241 AAIYLLTKTKLTLQTDSDGRFRIPGLCPDGKSLIKTTKVFAPIVLTMTKTSLKAAATIK 300  
 Db 241 AAIYLLTKTKLTLQTDSDGRFRIPGLCPDGKSLIKTTKVFAPIVLTMTKTSLKAAATIK 300

QY 301 AEFVRAETPYMVMNPETKARRAGOSVSLCCKATGKPRPDXYFWYHNDTLDDPSLYKHESK 360  
 301 AEFVRAETPYMVMNPETKARRAGOSVSLCCKATGKPRPDXYFWYHNDTLDDPSLYKHESK 360  
 Db 301 AEFVRAETPYMVMNPETKARRAGOSVSLCCKATGKPRPDXYFWYHNDTLDDPSLYKHESK 360

QY 361 LVLRKLDQOHAGEFYFCAQSDAGAVSKVAQLIYASDETPCNPVPESYLIRLPHDFQFN 420  
 361 LVLRKLDQOHAGEFYFCAQSDAGAVSKVAQLIYASDETPCNPVPESYLIRLPHDFQFN 420  
 Db 361 LVLRKLDQOHAGEFYFCAQSDAGAVSKVAQLIYASDETPCNPVPESYLIRLPHDFQFN 420

QY 421 ATNSFYVDVRCPEVTKCAGQDNGIRCRDAVONCCGISKTEEREIQCSTYLLPTKVAKEC 480  
 421 ATNSFYVDVRCPEVTKCAGQDNGIRCRDAVONCCGISKTEEREIQCSTYLLPTKVAKEC 480  
 Db 421 ATNSFYVDVRCPEVTKCAGQDNGIRCRDAVONCCGISKTEEREIQCSTYLLPTKVAKEC 480

QY 481 SCORCTETRSIVRGVSAADNGEPMRFGHYMGNSRVSMTGKOTFTLHVPQDTERLVLT 540  
 481 SCORCTETRSIVRGVSAADNGEPMRFGHYMGNSRVSMTGKOTFTLHVPQDTERLVLT 540  
 Db 481 SCORCTETRSIVRGVSAADNGEPMRFGHYMGNSRVSMTGKOTFTLHVPQDTERLVLT 540

QY 541 FVDRLOKFEVNTTVLPNNKGSANFHEIKMLRKEPTITLAMEFNITPLGEVGEDPMAE 600  
 541 FVDRLOKFEVNTTVLPNNKGSANFHEIKMLRKEPTITLAMEFNITPLGEVGEDPMAE 600  
 Db 541 FVDRLOKFEVNTTVLPNNKGSANFHEIKMLRKEPTITLAMEFNITPLGEVGEDPMAE 600

QY 601 LEIPSRSEFYRQNGEPYIGKVKASVTFELDPNNISTATAQTDLNFINDEGDTFFLRTYGMF 660  
 601 LEIPSRSEFYRQNGEPYIGKVKASVTFELDPNNISTATAQTDLNFINDEGDTFFLRTYGMF 660  
 Db 601 LEIPSRSEFYRQNGEPYIGKVKASVTFELDPNNISTATAQTDLNFINDEGDTFFLRTYGMF 660

QY 661 SVDFRDEVTSEPLNAGVKYVHLSDTVKMPBHISTVKLMSLNPDGTGLMEEGDFEFENOR 720  
 661 SVDFRDEVTSEPLNAGVKYVHLSDTVKMPBHISTVKLMSLNPDGTGLMEEGDFEFENOR 720  
 Db 661 SVDFRDEVTSEPLNAGVKYVHLSDTVKMPBHISTVKLMSLNPDGTGLMEEGDFEFENOR 720

QY 721 RNRREDTFLVGNLEIRERLFLNDVPESRCFVKVAYSERFLPSEOLOGVIVISYNL 780  
 721 RNRREDTFLVGNLEIRERLFLNDVPESRCFVKVAYSERFLPSEOLOGVIVISYNL 780  
 Db 721 RNRREDTFLVGNLEIRERLFLNDVPESRCFVKVAYSERFLPSEOLOGVIVISYNL 780

QY 781 EPRGTGLSNBRAMGRFSDVTGPNGACVPAFCDDQSPDAYSAVYLASLAGEELOAVESSP 840  
 781 EPRGTGLSNBRAMGRFSDVTGPNGACVPAFCDDQSPDAYSAVYLASLAGEELOAVESSP 840  
 Db 781 EPRGTGLSNBRAMGRFSDVTGPNGACVPAFCDDQSPDAYSAVYLASLAGEELOAVESSP 840

QY 841 KFPNNAIGVOPPLINKNTYRTRTHEDPRVKKTAQISMARPRPSAESSNGPIYAFENLR 900  
 841 KFPNNAIGVOPPLINKNTYRTRTHEDPRVKKTAQISMARPRPSAESSNGPIYAFENLR 900  
 Db 841 KFPNNAIGVOPPLINKNTYRTRTHEDPRVKKTAQISMARPRPSAESSNGPIYAFENLR 900

QY 901 ACEBAPSAHFRFYOLEGRYDNTVPFNEDDPMSTEDYLAWPMPMEERACIYIKVI 960  
 901 ACEBAPSAHFRFYOLEGRYDNTVPFNEDDPMSTEDYLAWPMPMEERACIYIKVI 960  
 Db 901 ACEBAPSAHFRFYOLEGRYDNTVPFNEDDPMSTEDYLAWPMPMEERACIYIKVI 960

QY 961 VGPLEVNVSRNMGTHRTVYGLKIRDYSTRDRDQPNVSAACLEFCKSGMLYDDDRY 1020  
 961 VGPLEVNVSRNMGTHRTVYGLKIRDYSTRDRDQPNVSAACLEFCKSGMLYDDDRY 1020  
 Db 961 VGPLEVNVSRNMGTHRTVYGLKIRDYSTRDRDQPNVSAACLEFCKSGMLYDDDRY 1020

QY 1021 DRTLKVIPIQSGCRASVNPMLHELYLVNHLPLAVNNDTSEYTMALADPLIGHNVGYTVT 1080  
 1021 DRTLKVIPIQSGCRASVNPMLHELYLVNHLPLAVNNDTSEYTMALADPLIGHNVGYTVT 1080  
 Db 1021 DRTLKVIPIQSGCRASVNPMLHELYLVNHLPLAVNNDTSEYTMALADPLIGHNVGYTVT 1080

QY 1081 DQDPRTAKEIALGRCFPGTSDGSSRIKSNVGAALFNCEYEROGVGRSAQYIQTSPAQS 1140  
 1081 DQDPRTAKEIALGRCFPGTSDGSSRIKSNVGAALFNCEYEROGVGRSAQYIQTSPAQS 1140  
 Db 1081 DQDPRTAKEIALGRCFPGTSDGSSRIKSNVGAALFNCEYEROGVGRSAQYIQTSPAQS 1140

QY 1141 PAAGTVGRVPSRROQASRGQSGVNASLRPPRAOOLIN 1184  
 1141 PAAGTVGRVPSRROQASRGQSGVNASLRPPRAOOLIN 1184  
 Db 1141 PAAGTVGRVPSRROQASRGQSGVNASLRPPRAOOLIN 1184

RESULT 2  
 US-08-996-083-3  
 ; Sequence 3, Application US/08996083A  
 ; Patent No. 6124095  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Magna, Holly  
 ; APPLICANT: Schaffer, Paul  
 ; APPLICANT: Lawton, Michael  
 ; APPLICANT: Yocum, Sue  
 ; APPLICANT: Mitchell, Peter  
 ; APPLICANT: Hutchinson, Nancy  
 ; APPLICANT: Murty, Lynn E.  
 ; TITLE OF INVENTION: HUMAN NUCLEOTIDE PYROPHOSPHOHYDROLASE-2  
 ; CURRENT APPLICATION NUMBER: US/08/996,083A  
 ; CURRENT FILING DATE: 1997-12-22  
 ; NUMBER OF SEQ ID NOS: 3  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 3  
 ; LENGTH: 1184  
 ; TYPE: PRF  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: misc.feature  
 ; OTHER INFORMATION: Incyte Clone No. 6124095: 422069  
 ; PUBLICATION INFORMATION:  
 US-08-996-083-3

Query Match 99.7%; Score 6297; DB 3; Length 1184;  
 Best Local Similarity 99.7%; Pred. No. 0;  
 Matches 1181; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MGTGKAVFSEFLVLEVTSLVGRQTMLTQSVRRVOPGKKNSIFAKPADTLESFGEWTTLF 60  
 1 MGTGKAVFSEFLVLEVTSLVGRQTMLTQSVRRVOPGKKNSIFAKPADTLESFGEWTTLF 60  
 Db 1 MGTGKAVFSEFLVLEVTSLVGRQTMLTQSVRRVOPGKKNSIFAKPADTLESFGEWTTLF 60

QY 61 NIDYGGKGYERLDAIRFYGGDVCARPLRLERATDWTDPAGSTGOVHGSFREGFWCL 120  
 61 NIDYGGKGYERLDAIRFYGGDVCARPLRLERATDWTDPAGSTGOVHGSFREGFWCL 120  
 Db 61 NIDYGGKGYERLDAIRFYGGDVCARPLRLERATDWTDPAGSTGOVHGSFREGFWCL 120

QY 121 NREORPQNSNTVTRFLCPGSLRRDTERIWSFWSKSCSAACGQTVQTRTRICLAE 180  
 121 NREORPQNSNTVTRFLCPGSLRRDTERIWSFWSKSCSAACGQTVQTRTRICLAE 180  
 Db 121 NREORPQNSNTVTRFLCPGSLRRDTERIWSFWSKSCSAACGQTVQTRTRICLAE 180

QY 181 MVSICSEASEGQHKMGQDCTACDLTCPMGOVNADCDACMCOFMLHGAVSLPGAPASG 240  
 181 MVSICSEASEGQHKMGQDCTACDLTCPMGOVNADCDACMCOFMLHGAVSLPGAPASG 240  
 Db 181 MVSICSEASEGQHKMGQDCTACDLTCPMGOVNADCDACMCOFMLHGAVSLPGAPASG 240

QY 241 AAIYLLTKTKLTLQTDSDGRFRIPGLCPDGKSLIKTTKVFAPIVLTMTKTSLKAAATIK 300  
 241 AAIYLLTKTKLTLQTDSDGRFRIPGLCPDGKSLIKTTKVFAPIVLTMTKTSLKAAATIK 300  
 Db 241 AAIYLLTKTKLTLQTDSDGRFRIPGLCPDGKSLIKTTKVFAPIVLTMTKTSLKAAATIK 300

QY 301 AEFVRAETPYMVMNPETKARRAGOSVSLCCKATGKPRPDXYFWYHNDTLDDPSLYKHESK 360  
 301 AEFVRAETPYMVMNPETKARRAGOSVSLCCKATGKPRPDXYFWYHNDTLDDPSLYKHESK 360  
 Db 301 AEFVRAETPYMVMNPETKARRAGOSVSLCCKATGKPRPDXYFWYHNDTLDDPSLYKHESK 360

QY 361 LVLRKLDQOHAGEFYFCAQSDAGAVSKVAQLIYASDETPCNPVPESYLIRLPHDFQFN 420  
 361 LVLRKLDQOHAGEFYFCAQSDAGAVSKVAQLIYASDETPCNPVPESYLIRLPHDFQFN 420  
 Db 361 LVLRKLDQOHAGEFYFCAQSDAGAVSKVAQLIYASDETPCNPVPESYLIRLPHDFQFN 420

```

421 AINSEFYDVGRCPVKTCAGQDNGIRCDAYONCCGISKEEREIOCSGYLLPTKVAKBC 480
422 AINSFYDVGRCPVKTCAGQDNGIRCDAYONCCGISKEEREIOCSGYLLPTKVAKBC 480
481 SCORCTETRSIVRGVSAADNGEPMRGHVYMGNSRVSMTGKGFPTLHVPODTERLVLT 540
482 SCORCTETRSIVRGVSAADNGEPMRGHVYMGNSRVSMTGKGFPTLHVPODTERLVLT 540
483 SCORCTETRSIVRGVSAADNGEPMRGHVYMGNSRVSMTGKGFPTLHVPODTERLVLT 540
541 FYDRLQKFPVNTKVLFPNNKSGAVFHEIKMLRKKEPITLEAMETNIIPGEEVGEPPAE 600
542 FYDRLQKFPVNTKVLFPNNKSGAVFHEIKMLRKKEPITLEAMETNIIPGEEVGEPPAE 600
543 FYDRLQKFPVNTKVLFPNNKSGAVFHEIKMLRKKEPITLEAMETNIIPGEEVGEPPAE 600
544 FYDRLQKFPVNTKVLFPNNKSGAVFHEIKMLRKKEPITLEAMETNIIPGEEVGEPPAE 600
601 LEIPRSRFRQNGEPYIGVKVASYTFILDPNRISTATATQDNLNINDEGDTPLRTYGMF 660
602 LEIPRSRFRQNGEPYIGVKVASYTFILDPNRISTATATQDNLNINDEGDTPLRTYGMF 660
603 LEIPRSRFRQNGEPYIGVKVASYTFILDPNRISTATATQDNLNINDEGDTPLRTYGMF 660
604 LEIPRSRFRQNGEPYIGVKVASYTFILDPNRISTATATQDNLNINDEGDTPLRTYGMF 660
661 SVDFRDEVTSEPLNAGKVYVHLDSQVKMPEHISTVKLMSLNDPTGLMEEGDFKEENOR 720
662 SVDFRDEVTSEPLNAGKVYVHLDSQVKMPEHISTVKLMSLNDPTGLMEEGDFKEENOR 720
663 SVDFRDEVTSEPLNAGKVYVHLDSQVKMPEHISTVKLMSLNDPTGLMEEGDFKEENOR 720
664 SVDFRDEVTSEPLNAGKVYVHLDSQVKMPEHISTVKLMSLNDPTGLMEEGDFKEENOR 720
721 RKRKRDRTFLVGNLEIRERLFLNDVPESRRCFVKVATRSERFLPSBOIQGVISVIL 780
722 RKRKRDRTFLVGNLEIRERLFLNDVPESRRCFVKVATRSERFLPSBOIQGVISVIL 780
723 RKRKRDRTFLVGNLEIRERLFLNDVPESRRCFVKVATRSERFLPSBOIQGVISVIL 780
724 RKRKRDRTFLVGNLEIRERLFLNDVPESRRCFVKVATRSERFLPSBOIQGVISVIL 780
781 EPRTFGLSNPRAMGRPDSVITGPNGACVAFCDOSPDAYSAYVLASLAGEELQAVESSP 840
782 EPRTFGLSNPRAMGRPDSVITGPNGACVAFCDOSPDAYSAYVLASLAGEELQAVESSP 840
783 EPRTFGLSNPRAMGRPDSVITGPNGACVAFCDOSPDAYSAYVLASLAGEELQAVESSP 840
784 EPRTFGLSNPRAMGRPDSVITGPNGACVAFCDOSPDAYSAYVLASLAGEELQAVESSP 840
841 KENPAIGVPODYLNKLNRRTDHEDPVKYKTAFOISMAKPRPNSAESNGPIYAFENLR 900
842 KENPAIGVPODYLNKLNRRTDHEDPVKYKTAFOISMAKPRPNSAESNGPIYAFENLR 900
843 KENPAIGVPODYLNKLNRRTDHEDPVKYKTAFOISMAKPRPNSAESNGPIYAFENLR 900
844 KENPAIGVPODYLNKLNRRTDHEDPVKYKTAFOISMAKPRPNSAESNGPIYAFENLR 900
901 ACEEAPPSAAHRRFYQIEEDRDYNTVPENEDDPMSTEDYLAAMPKPEHFRACITKVKI 960
902 ACEEAPPSAAHRRFYQIEEDRDYNTVPENEDDPMSTEDYLAAMPKPEHFRACITKVKI 960
903 ACEEAPPSAAHRRFYQIEEDRDYNTVPENEDDPMSTEDYLAAMPKPEHFRACITKVKI 960
904 ACEEAPPSAAHRRFYQIEEDRDYNTVPENEDDPMSTEDYLAAMPKPEHFRACITKVKI 960
961 VGPFLVNVSRMNGGTHRTYVGLKYGIRDVASTRDODPNVSAACLEFCSGLYDODVY 1020
962 VGPFLVNVSRMNGGTHRTYVGLKYGIRDVASTRDODPNVSAACLEFCSGLYDODVY 1020
963 VGPFLVNVSRMNGGTHRTYVGLKYGIRDVASTRDODPNVSAACLEFCSGLYDODVY 1020
964 VGPFLVNVSRMNGGTHRTYVGLKYGIRDVASTRDODPNVSAACLEFCSGLYDODVY 1020
1021 DFTLVKVIPOGSCSRASVNPMLHEYLNLPLAVNNDTSEYTMALPLDLGHNHYGYTYT 1080
1022 DFTLVKVIPOGSCSRASVNPMLHEYLNLPLAVNNDTSEYTMALPLDLGHNHYGYTYT 1080
1023 DFTLVKVIPOGSCSRASVNPMLHEYLNLPLAVNNDTSEYTMALPLDLGHNHYGYTYT 1080
1024 DFTLVKVIPOGSCSRASVNPMLHEYLNLPLAVNNDTSEYTMALPLDLGHNHYGYTYT 1080
1081 DODPTAKFIALGRCFDGTSDGSSRIMKSNVGVALTFCVERQVGRSOFYLOQSTPAOS 1140
1082 DODPTAKFIALGRCFDGTSDGSSRIMKSNVGVALTFCVERQVGRSOFYLOQSTPAOS 1140
1083 DODPTAKFIALGRCFDGTSDGSSRIMKSNVGVALTFCVERQVGRSOFYLOQSTPAOS 1140
1084 DODPTAKFIALGRCFDGTSDGSSRIMKSNVGVALTFCVERQVGRSOFYLOQSTPAOS 1140
1141 PAAGTVQGRVPSRRQORASRGQORSGVVASLRFPRVAAOOLPIN 1184
1142 PAAGTVQGRVPSRRQORASRGQORSGVVASLRFPRVAAOOLPIN 1184
1143 PAAGTVQGRVPSRRQORASRGQORSGVVASLRFPRVAAOOLPIN 1184
1144 PAAGTVQGRVPSRRQORASRGQORSGVVASLRFPRVAAOOLPIN 1184
```

```

RESULT 3
US-08-996-083-1
Sequence 1, Application US/08996083A
```

```

GENERAL INFORMATION:
Patent No. 6124095
```

```

APPLICANT: Magna, Holly
APPLICANT: Schaffer, Paul
APPLICANT: Lawton, Michael
APPLICANT: Yocum, Sue
APPLICANT: Mitchell, Peter
APPLICANT: Hutchinson, Nancy
```

```

TITLE OF INVENTION: HUMAN NUCLEOTIDE PYROPHOSPHOHYDROLASE-2
```

```

FILE REFERENCE: PF-0420 US
```

```

CURRENT APPLICATION NUMBER: US/08/996, 083A
```

```

NUMBER OF SEQ ID NOS: 3
```

```

SOFTWARE: FastSeq for Windows Version 3.0
```

```

; SEQ ID NO 1
```

```

; LENGTH: 1156
; TYPE: PR
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc-feature
; OTHER INFORMATION: Incyte Clone No. 6124095; 1388013
; PUBLICATION INFORMATION:
US-08-996-083-1
```

```

Query Match 48.9%; Score 3091.5; DB 3; Length 1156;
Best Local Similarity 50.3%; Pred. No. 5,3e-275;
Matches 584; Conservative 178; Mismatches 359; Indels 39; Gaps 12;
```

```

24 TMLTQSVRVQGGKKNPSIFAKPADTLESBGEWTTLFNIDYPOGKDYERLDAIRTYGD 83
32 TALGLERRSVYTGQPSPAL-----EDMEASEWTSWFNVDHDPGDDDFSLAIRPYG 86
84 -RYCARPLRLAARTTMTWTSAGTVGVHGSPRGFCMLNREORPGONCSNTRYFLCPCPG 142
87 ARVCPRLLEARTTDMALPSAVGERVHLNPTRGFCMLNREORPGRCNSYHVRFCP-- 144
143 SLRDRTERIWSPPWSKSCSAACGOTGVOTRRICLAENVSLCSEASEGOMCMGDCYTA 202
145 -----LEASMGAMGPMWPGSGSG- -GRRLLRRHCHPSPGADCPGRPLLEAKVRRPCG 198
203 CDLTGCMGOVNADCDKCMODEFHLGAVSLPGAPASGAIIYLLTPKLLTQDSDGRF 262
199 CSL-----DICEPDDHLLGSVYTPSQDPLLGARVSLRD- -PGTVATSDAHGTF 246
263 RLPGLCPDGKSLIKTKYKF-APIVLTPKTSLSKATTAEFVRAETPYVMNPETKARR 321
247 RVPGVCAADRANIRAOQMDGFSAGEAOANGSISVYTTILD- -KLEKPLVYKHRESVRE 304
322 AGQSVSLCCATGKPRPRDKFYVYHNDTLLDPSLYKHESKLYLKKLOHQHAGEYFCAQSD 381
305 AGQNTVECCKASGTPMKRYSMFNHGTLLDRAHNGGAHLELRGLRPDQAGIYHCKAMNE 364
382 AGAVKSKVAVQLYVTSADPEPCNPVPESYILRLPHCFQONATNSFFYDVGRCPVKTCAGQ 441
365 AGAVRSGTARLTVLAPGQCDPRPREYLIKLPEDCGQSGSGAYLDVGICPTRPCPSLA 424
442 DNGIRCDVONCCGISKEEREIOCSGYLLPTKVAKESCORCTETRSIVRGVSAAD 501
425 GSSPRCGDASSRCCSVRLERREIHCPGYVLPKYVAECGQCLPRLGLVGRVAAAS 484
502 GEMRGRGVYMGNSRVSMTGKGFPTLHVPODTERLVLTFFVDRLOKFEVNTTKVLPNNKG 561
485 GEPLRFARILLGOEPIGFAYOGDFTIEVPPSTQRLVYTFVDPSSGEMDAVRLPDPDGG 544
562 SAVFHEIKMLRKKEPITLEAMETNIIPGEEVGEEDPMALETPRSFYQNGEPYIGKVK 621
545 AGVYHEVKAMRKKAIVILHTSOSNTIPLGELEDEALGELVLPFSGAFRADGPKYSGPVE 604
622 ASVTFLDPNRISTATATQDNLNINDEGDTPLRTYGMFVSVPFRDVTSEPLNAGKVYKH 681
605 ARVTEVDPDRDLTSASASAPDLPVSDGELAPRTYGMFVSVDLRADGSAEQLOVGVAAVR 664
682 LDSTQVKMPEHISTVKLMSLNDPTGLMEEGDFKEENOR- -RKRKRDRTFLVGNLEIRER 739
665 VASQIHPGHEALKLMSLNPETGLMEESGRRRGSSGPRVRRERFVLGVNVEIRER 724
740 RLFLNDVPESRRCFVKVATRSERFLPSBOIQGVISVYNLEPRGFLSNPRAMGRFDSV 799
725 RLFLNDVPESRRCFVKVATRSERFLPSBOIQGVISVYNLEPRGFLSNPRAMGRFDSV 784
800 ITGPNACVAPCDOSPDAYSAYVLASLAGEELQAVESSPKRNPNAITGVPODYLLKLVY 859
785 VTGPNACVAPCDADRPATVATLATTGEELEAPSLPRPLPVTGVTOCTYDLRLGT 844
860 RTDHEDPVKYKTAFOISMAKPRPNSAESNGPIYAFENLRACEAPPSAAHRRFYQIEG 919
845 RTDHDPAFAKRNGRFINLAKPRPDGPADANGVYVWRSRLREGOGAPVYASHRRFAVEA 904
```

```

0Y 920 DRYDNYNPFEDDDPMSTEDBYLAMPKPMERACVTKVIVPLEVANSRMGSTRR 978
Db 905 DKREYNVPFREDGPASTVDGLAMWNPQDEFACFLAKXIGQPQCYMVRSMNAGSGHR 964
0Y 960 TVKLXLYGRDVSRTRDRDQPNVSAACELEFGCSMLYDQDRVDFTLVKVIPQSCRRASVN 1033
Db 965 TRGOIYGLRDARSVRDEPERPTSAAECFEFGCSMLEFDQRQVDRFTLVITIMPOGSCRVAVN 1024
0Y 1040 PMLHEYLNVNHLPLAVNNDTSEVTMLAPLDPLGNHYGYTPTODDPRTAKEIALGRCFOST 1099
Db 1025 GILRLYLRHPPYPADBPFAFSMLAPDLGHNYGVYTSDPSPLAKEIALGRCFDS 1088
0Y 1100 SDGSSRIKSNVGVALTENCVERGROSAFOYLQSTPAQSPACTVOGVRPVSRRQARAS 1155
Db 1085 SDGFSEKKADAGTAATVFQCCEHPAGRPRLFQLLESFA--TALGDIR-----REMSEA 1133
0Y 1160 RGGOROSCVASLRFPRYAQ 1179
Db 1138 QAOARASGPLRTRR-GRVRQ 1156

UNIT 4
US-09-429-516-1
; Sequence 1, Application US/09429516
; Patent No. 6251389
; GENERAL INFORMATION:
; APPLICANT: Magna, Holly
; APPLICANT: Schaffer, Paul
; APPLICANT: Lawton, Michael
; APPLICANT: Vocum, Sue
; APPLICANT: Mitchell, Peter
; APPLICANT: Hutchinson, Nancy
; APPLICANT: Murry, Lynn E.
; TITLE OF INVENTION: HUMAN NUCLEOTIDE PYROPHOS
; TITLE OF INVENTION: PHOHDROLASE-2
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/429, 516
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/996,083
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0420 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1156 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: SATPFI1008
; CLONE: 1388013,

```

US-09-429-516-1

Query Match	48.9%;	Score 3091.5;	DB 4;	Length 1156;
Best Local Similarity	50.3%;	Pred. No. 5.3e-275;		
Matches 584;	Conservative 178;	Mismatches 359;	Indels 39;	Gaps 12.
24 TMLTOSVRRVYQGGKKNPSIFAKPADLTLESPPMTTLFNIDYGGKGDYRLDAIRFYIGD	83			
32 TALGLERSVYVYGGQSPAL-----EDWEASMTSMFNVDHFGGQDPEFSLAIRFYVGP	86			
84 -RYCAPRLFLKRTDMPAGSTGOVYHGSPPDEGMCLNREORPGQCNSTYVRELCPGG	142			
87 AAVCPRPFLAKRTDMLPVSAGEVHMLNPTRGFMCLNREORPRRCRCNHYVRRCP--	144			
143 SLRPRDRIWSPWSPMSKCSAACGGQGVOTPRICLAEWYSLCSEASEGGQCMQODCTA	202			
145 -----LEASWGMGMGPGPCSGCG-PGRRLRRRHCHSPAGADCPGRFLPAQCVRRPCG	198			
203 CDLTCPMGVWADCDACMCQDFMLHGAVSLPGAPASAGAIYLLTKPKLLTQTDSDGRF	262			
199 CSL-----DICEPDRHLLGSSVYTPSGQRPILGARVSLRQ-PQTVATSDAHGTF	246			
263 RIRPGCDPSKSLIKTKYKF-APYVLTMKTKSLKATITAEFVRAATPRVMNPPFKARR	321			
247 RVPQCADSRANIRQMDGFSAGEAOQANGSTVYTLID--KLEKPLVNHPSRRVPE	304			
322 AGOSVSLCCATGATGPRDXYFYHNDTLDPDSLKHESKLVLRKLQDQAGEYFCKAOSD	381			
305 AGQNTVFCCKASGTPMKKYSWFNHNGTLLDRRAHGVAHLELRGLRPDQAGIYHCKANE	364			
382 AGAVSKVAQLLYTASDEFPCNPVPESTYILRLPHCFQNAITSFYDYVRCRYKTCAGQO	441			
365 AGAVRSGTARLVYLAQGPADPRREYLIKLPEDCGQSGSPALVDLCPDTRCPSTA	424			
442 DNGIRDADVQACCGISKREERIOCSGYTLTPTKVAKESCSQCNCTETRSIVGRVSAADN	501			
425 GSPRCGDASSRCCSVRRLERRREIHGQVLPVKVAVECGCQKCLPRGLVGRVAAUDS	484			
502 GEPNRFAGHYMGNSRYKSMTGKGYFTLHVPQDTERLYTLFVDRLQKFVYTKVLPNNKG	561			
485 GEPLEFRARILLQEPGIFGAYOGDFTEIYVPSQRIQVLFVTEVSGEFGMAVRLPDPGG	544			
562 SAVFHEIKMLRKREPTLTAMETNIIPLAEVVGEDMALETIPRSFRYRONGEPYIGKYK	621			
545 AGYVHEHYKAKMKRKAPVILTLTSSQNTIPLQELDEARLIGELVLPQSGAFRAADCKYGPVE	604			
605 ARVTEFVDPDLTISASAPSDLRFVSDGDELAPLRTYGMFSVDIRAFAGSAEOLQVGFVAAR	664			
682 LDSTQKMEHEHSTVYLMWLNPDTPGMEEGDPRKFNQO--RRKREDRPRFLVGNLEIRER	739			
665 VVASQJHMHGHEVALKWLMLNPEPTGLMEESGGRREGSGSPRYRRERERVLGNVETIRER	724			
740 RLFLNDLPESRRCFYVRAVYRSERLPLSDQIQQVYISVNLBPTGFLSLNPRAMGRHDSV	799			
725 RLFLNDLPERRRCFYVRAVYANDKFTPSQVQVEGVVTVLNLBPAPGFSANPRAMGRFDSA	784			
800 ITGPNACYPARCDDSPRAYSYVYLAISAGELQAVESSPKFNPAITGVPQRYLKNLKY	859			
785 VYGPNGACLPAPCDARPRAPYATLVATLAGEBELPAPSLPRPLPVTAGTQPYLDRLQY	844			
860 RRTDHEDDPVYKTAFAFISNAKPPNPSAEESNGITYAFEMLRACEEAPSAHFREVOIG	919			
845 KRDDHDDPFAKRRGFTINLAKPRPGDPAEANGFYVYWRSLRECQGPATYASHFRFAKVA	904			
920 DRYDNTVTFENEDDPKSMTEEDYLAMPKMEEFKACYIKVLYGPLEVYNSRMMGSHRR	979			
905 DKYEYVNVFRECTPASMTGDLILAMPNPQEFERACFLAKYIQGQPYWYRSHNAGSGSHR	964			
980 YNGKYGIRDVNSTRRDQPNVSAACLEKRGSMALYDQDRVDTLYKVLPOGSCRASVYN	1039			

Db 965 TRGOLYGLRDARSVDPERPCTSAACVEFKCSGMLFDOROVDRTLVLTIMPOGSCRRVAVN 1024  
Qy 1040 PMLHELYLVNHLPLAVNNSTSETMLAPDLPGHNGIYVTDODPRFAKETAIGRCFCGT 1099  
Db 1025 GLRDYLTFRHPPVPAEDPAFASMLAPDLPGHNGIYVTDQSRKLAKETAIGRCFCGS 1084  
Qy 1100 SDGSSRIKSNVGVALTENCVERQVROSARFOYLOSTPAQSPAACTVOGRVPSRQORAS 1159  
Db 1085 SDGSRREKADAGTAVTFQCCEPPAGRPSLFORLLESFA--TALGDIR-----REMSBA 1137  
Qy 1160 RGGORSGVVASLRRPVAQ 1179  
Db 1138 QAOARASGRLTRR-GRVQ 1156

RESULT 5  
US-09-429-516-3  
Sequence 3, Application US/09429516  
Patent No. 6251389  
GENERAL INFORMATION:  
APPLICANT: Magna, Holly  
APPLICANT: Schafter, Paul  
APPLICANT: Lawton, Michael  
APPLICANT: Yocum, Sue  
APPLICANT: Mitchell, Peter  
APPLICANT: Hutchinson, Nancy  
APPLICANT: Murry, Lynn E.  
TITLE OF INVENTION: HUMAN NUCLEOTIDE PYROPHOS  
NUMBER OF INVENTION: PHOYDROLASE-2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/429,516  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/996,083  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0420 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-855-0555  
TELEFAX: 650-845-4166  
TELEX:  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1156 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: SATF1002  
CLONE: 422069  
US-09-429-516-3

Query Match 48.98; Score 3091.5; DB 4; Length 1156;  
Best Local Similarity 50.38; Pred. NO. 5.3e-275;  
Matches 584; Conservative 178; Mismatches 359; Indels 39; Gaps 12;

Qy 24 TMLTQSVRRVOPGKKNPISIFAKPADTLESPEGEMTTLFINDYPGKGKDYERLDAIRFYCD 83  
Db 32 TALGERSVYVTGQSPRL-----EDMEEASGEMTWFVNDHGGGDESLAIFHYIGP 86  
Qy 84 -RVCARPLREARTTDMTPAGSTGOVHGSPEGFCMLNREORPGQNSNTTVRFCLPQG 142  
Db 87 ARVCRPLALEARTTDMALPSAVGERVHLPTRGFWCLNREQPRGRCSNHYHFRCP-- 144  
Qy 143 SLRRTERTNWSWMSWMSCSAAGCTGYOTRTRICLAEMVSLCSASREGCHMKQDCTA 202  
Db 145 -----LEASWGAAGPWCSCSCG-PGRRLRRRHCPSPAGACPRPLEACKCVRRPCG 198  
Qy 203 CDITCPMGQVANAADCACACQOPFMLHGAVALPGAPASCAAIYLLTKTKLLTQTSDFR 262  
Db 199 CSL-----DTCPCPHIILGSVYTPSGOPLDARVSLRQ-PGYATSDAHGTF 246  
Qy 263 RIPGLCPDGKSLIKITTKVF-APIVLTMPKTSLKATTKAEFVRAETPYVMNPETKARR 321  
Db 247 RVPVCADSRANIRAQMDGFSAGEAQAQANGSISVVTILD-KLEKPYLVKHPBSRYRE 304  
Qy 322 AGQSVSLCCKATGKRPDKTYFVHNDTLLDPSLYKHESKLVLRKIQHQAQGEYFCKAQSD 381  
Db 305 AGQNTVFCCKASGTPMPKYSWFMNGTLLDRRAHGYAHLEIRLPDQAQGIYHCKANNE 364  
Qy 382 AGAVSKVAOLIVTASDETPCNPVPESTLIRLPHDCFQMNATNSFYDYDGRCPVKCAQO 441  
Db 365 AGAVASGTARLTVLAPGCPACDPREYILKLPEDCCGPGSPAYLDGLDPTDRCPSLA 424  
Qy 442 DNGIRCDVAVNCCGISKTEREIOCSGYTLPTVKAKECSQCRCTETRSIVRGVSAADN 501  
Db 425 GSPRCGDASSRCSVRRLREIRHCPGYLVKVAACGCGKCLPRGLVRGRVAAADS 484  
Qy 502 GEPARFGVYVNGSNRVSMTGKGTTLVPODTEVLVTFVDRLOKFNVTIKVLPFNKG 561  
Db 485 GEPLRFARILIGQEPDIGTAVOGDETLVFPDPSGEFADAVRVLVFPDRG 544  
Qy 562 SAVFEIKMLRKEPITTEAMETNIIPGEVYGEDPMALFIPSRFONGCEPIYGK 621  
Db 545 AGVYHEVAMRKKAIVILHTSOSNTIPLGELEDEAPLELVLPSCAFRADCKPISGVE 604  
Qy 605 ARVTFVDRDLTASASASDRLFVDSDELAPLRTYGFMSVDLRAFGAEOLOVGPVAVR 664  
Qy 622 ASVTFPLDRNISTATAVTAOTDLNFINDEGDTEPLRTYGFMSVDFRDEVTSEPLNACKVYH 681  
Db 605 ARVTFVDRDLTASASASDRLFVDSDELAPLRTYGFMSVDLRAFGAEOLOVGPVAVR 664  
Qy 682 LDSTOVKAPERHISTVYKLSLNDPTGLMEBGDFEKNOR--RNRKEDRTFLVGLNETIER 739  
Db 665 VAAEQIHMFGVHEALKLSLNPETGLMEESGFRREGSSGPRVREERVFVLTGANEIRER 724  
Qy 740 RLENLDVPESSRCFVKVAVYSEERFLPSEOGVYISYINLEPRKGYLNPBRAMGRPSV 799  
Db 725 RLFLNDVPESSRCFVKVAVYANDKTPSEOVEGVYVTLVNLPEAPGFSANBRAMGRPSA 784  
Qy 800 ITGPNAGVPAFCDDQSPDAYSAVYVLAAGELOAVESSPKFNMAIGVPOPYLNKLN 859  
Db 785 VTGPNAGVPAFCDDQSPDAYSAVYVLAAGELOAVESSPKFNMAIGVPOPYLNKLN 859  
Qy 860 RTHEDPRVYKTAFOISMARPRNSAEBSNCPITAFENLRACEAPSAHFRYQLEG 919  
Db 845 RTHEDPRVYKTAFOISMARPRNSAEBSNCPITAFENLRACEAPSAHFRYQLEG 919  
Qy 920 DRYDNTVFPFNDDPMSTEDVLIAMWPKPMEFRACIYKIVGPLEVAVNRSHNMGTFR 979  
Db 905 DKLEYNVYFPFRGCTPASTGTGLIAMWPKPMEFRACIYKIVGPLEVAVNRSHNMGTFR 964  
Qy 980 TVGKLYGIRDVSTRDRQDPNVSAACLEFKCSGMLYDODRVDRTLVKVYIPGSCRRASVN 1039  
Db 965 TRGOLYGLRDARSVDPERPCTSAACVEFKCSGMLFDOROVDRTLVLTIMPOGSCRRVAVN 1024  
Qy 1040 PMLHELYLVNHLPLAVNNSTSETMLAPDLPGHNGIYVTDODPRFAKETAIGRCFCGT 1099  
Db 1025 GLRDYLTFRHPPVPAEDPAFASMLAPDLPGHNGIYVTDQSRKLAKETAIGRCFCGS 1084  
Qy 1100 SDGSSRIKSNVGVALTENCVERQVROSARFOYLOSTPAQSPAACTVOGRVPSRQORAS 1159

DB 1085 SDFSTHEMADATATVTPQCRPEPPAGRPSLFOFLLESPPA--TALGDIR-----REMSEMA 1137  
 QY 1160 RGSOROSGVASLRFPRVAO 1179  
 DB 1138 QAQARASGLPTRRR-GRVRO 1156

## RESULT 6

US-08-954-333-7  
 ; Sequence 7, Application US/08954333  
 ; Patent No. 5986080  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Ikuko Masuda, et al.  
 ; TITLE OF INVENTION: CLONED NUCLEOTIDE PYROPHOSPHOHYDROLASE AND  
 ; TITLE OF INVENTION: USED THEREOF  
 ; NUMBER OF SEQUENCES: 9  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Quaries & Brady  
 ; STREET: 411 East Wisconsin Avenue  
 ; CITY: Milwaukee  
 ; STATE: Wisconsin  
 ; COUNTRY: U.S.A.  
 ; ZIP: 53202-4497  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patentin Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/954,333  
 ; FILING DATE:  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Ryser, David G.  
 ; REGISTRATION NUMBER: 36,407  
 ; REFERENCE/DOCKET NUMBER: 650053, 91070  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (414) 277-5717  
 ; TELEFAX: (414) 271-3552  
 ; INFORMATION FOR SEQ ID NO: 7:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 599 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: peptide  
 ; US-08-954-333-7

Query Match 45.9%; Score 2898.5; DB 2; Length 599;  
 Best Local Similarity 91.1%; Pred. No. 9,6e-258;  
 Matches 544; Conservative 23; Mismatches 29; Indels 1; Gaps 1;

QY 588 PLEGVVGEDPMALIEIPSRFSYRQNGEPYIGVKASVTFPLDPNNISTATTAOTDLNFIND 647  
 DB 4 PLEDMGDEPMGLIEIPSKFSYRQNGEPYIGVKASVTFPLDPNNISTATTAOSDINFIND 63  
 QY 648 EGGTFPLRTYGMFSVPFREVSEPIINACKVKVHLDSQVKKMPEHISTKLSINPDTGL 707  
 DB 64 EGGTFPLRTYGMFSVPFREVSEPIINACKVKVHLDSQVKKMPEHISTKLSINPDTGL 123  
 QY 708 WEDEGDFKFNORRNKREDRTFLVNLIERRLFNLDVPESSRCEVKKVAYRSEFPLDS 767  
 DB 124 WEDEGDFKFNORRNKREDRTFLVNLIERRLFNLDVPESSRCEVKKVAYRSEFPLDS 183  
 QY 768 EOIQGVIVSINLEPRTGFLSNRANGRPDSVITGNGACVPAFCDDOSPDASAVVLAS 827  
 DB 184 EOIQGVIVSINLEPRTGFLSNRANGRPDSVITGNGACVPAFCDDOSPDASAVVLAS 243  
 QY 828 LAGEELQAVSSPFENNAIGVPOPYLNLKLNERTDHEDPVKKTAFOISMAKPRNSAE 887  
 DB 244 LAGEELQAVSSPFENNAIGVPOPYLNLKLNERTDHEDPVKKTAFOISMAKPRNSAE 303

QY 888 ESNQPIYAFENLRACEAPPSAAHPRFYQIEGDRDYDNTVPENEDDPMSTEDYLAMMPK 947  
 DB 304 ESNQPIYAFENLRACEAPPSAAHPRFYQIEGDRDYDNTVPENEDDPMSTEDYLAMMPK 363  
 QY 948 PMEFKACIKYKIVGPLEVNVSRNMGSTRRTVYKGLYGRDVRSTDRDOPNVSAACLE 1007  
 DB 364 PMEFKACIKYKIVGPLEVNVSRNMGSTRRTVYKGLYGRDVRSTDRDOPNVSAACLE 423  
 QY 1008 FPCSGMLDQDRVDRTLVKVIPOGSCRRASVNPMLHEVLVNLPLAVNNDSEYTMALAPL 1067  
 DB 424 FPCSGMLDQDRVDRTLVKVIPOGSCRRASVNPMLHEVLVNLPLAVNNDSEYTMALAPL 483  
 QY 1068 DPLGHNYGIYTVTDQDPRTAKEIALGRCEGTSDDGSRIRKMSNVGVALTFNCVEROYGRQ 1127  
 DB 484 DPLGHNYGIYTVTDQDPRTAKEIALGRCEGTSDDGSRIRKMSNVGVALTFNCVEROYGRQ 543  
 QY 1128 SAFQYLOSTPASPAGTYQGRVPSRROORASRGORGVASLRFPRVAOQPLIN 1184  
 DB 544 SAFQYLOSTPASPAGTYQGRVPSRROORASRGORGVASLRFPRVAOQPLIN 599

## RESULT 7

US-08-313-288B-20  
 ; Sequence 20, Application US/08313288B  
 ; Patent No. 5750502  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Jessell, Thomas M. and Avihu Klar  
 ; TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A  
 ; TITLE OF INVENTION: NOVEL SECRETED PROTEIN, F-SPONDIN  
 ; NUMBER OF SEQUENCES: 20  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Cooper & Dunham LLP  
 ; STREET: 1185 Avenue of the Americas  
 ; CITY: New York  
 ; STATE: New York  
 ; COUNTRY: USA  
 ; ZIP: 10036  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patentin Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/313,288B  
 ; FILING DATE: January 5, 1995  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: White, John P.  
 ; REGISTRATION NUMBER: 28,678  
 ; REFERENCE/DOCKET NUMBER: 40028-A-PCT-US  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (212) 278-0400  
 ; TELEFAX: (212) 391-0526  
 ; INFO:  
 ; INFORMATION FOR SEQ ID NO: 20:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1170 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: peptide  
 ; US-08-313-288B-20

Query Match 2.5%; Score 157; DB 1; Length 1170;  
 Best Local Similarity 20.1%; Pred. No. 5.2e-05;  
 Matches 111; Conservative 37; Mismatches 150; Indels 254; Gaps 27;

QY 10 SFLVEVTSVLGRQMLQSVARVOPGKKNPSIFAKPADTLESF-----GEW 56  
 DB 278 SSMVLELRRLRTIVTTLDLSIKVTEENK-----ELANEILRRPRLCYHNGVQYRNDEW 331

```

OY 57 TT-----LFNIDYPGKGDYERLDAIRFYXGDRVCARPLRL 92
DB 332 TVDSCTECHCQSVYICKKVCSPIMPCSNATVPDGE-----CCPRCWS 375
OY 93 EARTDWTTPAGSTGOVYHGSPPREGFW-----CLNREORPGONCSNYTVRFLCPGSLRL 146
DB 376 DSADGWSWSPWSE-----WTSCSTSCGNGIQGRGSCDILNRR--CEGSSVQT 420
OY 147 DHERI-----WSPWSPWKSCKSAACGOTGVOTRTRICLAEMVSL-----CSEASE 190
DB 421 RYCHIQECCDKRKQDGSWSHSPWSSCVTCGD--GVITRIRLCNSPSPOMNKPCEGEAR 479
OY 191 ESOHCMGODC-----TACDLTC-----PMGOVNA--DC-----D 217
DB 480 ETKACKKADACPIINGGMPSPWDICSVTCGGVOKRSRLCNPTQPGKDCVGVDTENO 539
OY 218 ACMODFMLHGAVALPGCAPASGAAYLLTKYPKLLTQDSDGRPRIPGLCPGKSLIKI 277
DB 540 ICKKDCPIDGLSNPCFAGVK-----CTSTP-----GSMWC--GACPPGYS----- 580
OY 278 TVKRAPIVLTMPKTSIKATIKAEFVRAETPYMMNPETKARRAGQSVSLCCKATGRPR 337
DB 581 -----GNGIQCTDVECKEV 595
OY 338 PDKYFWYHNDLLDPSLYKHESKLYLRKLOHQAGEYFCFKAQSDAG----- 383
DB 596 PDACE--NNH-----GEHRE--NTDPOYCNLPCPPRTGSO 628
OY 384 ----AVKSVAOIIVTASDETPCNVPESYLRIRPHDGFONATNSF---YVD--VGRCPVK 435
DB 629 PFGQVEVHAHTAKQVC---KPRNCTDG---THDCNKNAKVCYLI GHYSPMYRCECK 679
OY 436 TCAGOODNGIRC 447
DB 680 --PGTAGNGITC 689

```

## RESULT 8

US-08-918-914-4

Sequence 4, Application US/08918914

Patent No. 5876963

GENERAL INFORMATION:

APPLICANT: Mitchell, Peter

APPLICANT: Hutchinson, Nancy

APPLICANT: Lawton, Michael

APPLICANT: Magna, Holly

APPLICANT: Murry, Lynn E.

TITLE OF INVENTION: HUMAN NUCLEOTIDE PYROPHOSPHORYLASE

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESSEE: Incyte Pharmaceuticals, Inc.

STREET: 3174 Porter Dr.

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/918,914

FILING DATE: Filed Herewith

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J.

REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: PF-0369

TELECOMMUNICATION INFORMATION:

```

TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 788 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 1070094
US-08-918-914-4

```

Query Match 2.3%; Score 147.5; DB 2; Length 788;  
 Best Local Similarity 27.2%; Pred. No. 0.00019;  
 Matches 55; Conservative 15; Mismatches 65; Indels 67; Gaps 11;

```

OY 46 PADTLESPEWMTTLFNIDYPGKGDYERLDAIRF-----YYGDRVCARPLRL 93
DB 501 PCQTFSEWCWST---CSASCGSGORER---TRFCHLGTNRCEGKDYESQCSAGCPBW 554
OY 94 ARTFTWTPAGST--GOVYHGSPPREGFWCL-----NREOR--PGONCSNYTVRFL 138
DB 555 SQWEDMGQCVTCGGGVAVRQRT---CLGGVFGDHLGQGPKEGRACDGGPCS----- 604
OY 139 CPPGSLRDTERTIWSWSPWKSCKSAACGOTGVOTRTRICLAEMVSLCSASSEGOHCWQ 198
DB 605 -----LWSPWQSMSTCSASCG--SGMKRRQVC--QFGIDCGPNEESQFCYGP 649
OY 199 DC-----TACDLTCPMQ 211
DB 650 PCAEWTEMCWMSGCSSKCGPGQ 671

```

## RESULT 9

US-08-348-006B-5

Sequence 5, Application US/08348006B

Patent No. 5658756

GENERAL INFORMATION:

APPLICANT: RODAN, GIDEON A.

APPLICANT: SCHMIDT, AZRIEL

APPLICANT: RUTLEDGE, SU JANE

TITLE OF INVENTION: CDNA ENCODING A NOVEL HUMAN PROTEIN

NUMBER OF SEQUENCES: 7

CORRESPONDENCE ADDRESS:

ADDRESSEE: J. MARK HAND

STREET: 126 E. LINCOLN AVE., P.O. BOX 2000

CITY: RAHWAY

STATE: NEW JERSEY

COUNTRY: USA

ZIP: 07065-0900

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC Compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/348,006B

FILING DATE:

CLASSIFICATION: 514

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/122,032

FILING DATE: 14-SEP-1993

ATTORNEY/AGENT INFORMATION:

NAME: HAND, J. MARK

REGISTRATION NUMBER: 36,545

REFERENCE/DOCKET NUMBER: 189921A

TELECOMMUNICATION INFORMATION:

TELEPHONE: 908-594-3905

TELEFAX: 908-594-4720

; INFORMATION FOR SEQ ID NO: 5:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1911 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-08-348-006B-5

Query Match 2.3%; Score 144; DB 1; Length 1911;  
 Best Local Similarity 19.5%; Pred. No. 0.0019;  
 Matches 186; Conservative 122; Mismatches 362; Indels 284; Gaps 46;

QY 295 KAATIKAEFVAETPYVWNPETARRAGOSVSLCKATGKPRDKTFWYHNDLLDPSL 354  
 DB 121 KLTVLREQLPSGFPNIDMGPKLVKERTATMLCAASGNPDE-TWTFDPLVPDPSA 179  
 QY 355 YKHESK-----LVLRKIQHQAGEYFCKAOSDAGAVSKVAQLIVTASDETPCNPDES 408  
 DB 180 SNGRIKQRLSGALQIESSEETDOGKECVATNSAGVRYSSANLYRVYR-----RVAPRF 234  
 QY 409 YLIRLPHDCFQNA-TNSFYDYGR-CP-VKTCAGQDNGICRDVAVONCCGISTEEREI 465  
 DB 235 SILPMSHIMPGANNITICVAVGSPMPYKMMQGAED-----LTPEDDM 278  
 QY 466 QCSGYTLP-TKYAKECSCQCRTESTRSYVGRVSAADNGEPMRFGHYVGNRSVMTGY-K 523  
 DB 279 PVGRNVLELTVDKDSANHPCVAMSSL--GVIEAV-----AQITVKSLEPK 321  
 QY 524 GFTFLHVPQDTERLVLTIVYDRLQKFNVTTKVLPEKKGSAAVHEIKMLRKEPTILEAME 583  
 DB 322 APGTMTVENTATSTITTWDSGNPDVSVYVIEYKSKQDGPYQK-----EDIT 371  
 QY 584 TNIPLGVEVDEPMALLET-----PSRSFYRQNGEPIYGVKASVTFPLDRN 631  
 DB 372 TTRYSIG---GISPSSEYEIWSAVNSIGQRPSPESVYTRTGEQAPAR-----PPRN 420  
 QY 632 ISTATAQTDNFINDEGDTPLRTYGMFSVDERGEVSEPTLNAGKVHLDSFOVKME 691  
 DB 421 VOARMLASTTKLVQWEE---PVEPNGLIR-GYRYTYME-----PE 457  
 QY 692 HISTVKLMSLNDTGLMEEGDFKFEENQRNKRREDRTFLVGNLEIRERLENLDVPESRR 751  
 DB 458 H-----PVGNMW-----QKHNYDSLLTIVGSL-----LEDET 484  
 QY 752 CTVKVRAYRSEFLP-SEQI-----QGVVISYINL--EPRGFLSNPRAMG--RFDVYL- 800  
 DB 485 YTVRVIAFTFSGDGLSDPTQVKTQOGVPGOPMNLRAEKSE-TSTITLWSPPROESTIIL 543  
 QY 801 -----TGPNGACVPAFCDDQSPDAYSAVVLASLAGEELQAVESSPKENPNAIGVPOYL 854  
 DB 544 YELLREGDGHREV-----GRTFDPTTSYVVEDLKPNTEYAFRLAR--SPQGLAFTPVY 597  
 QY 855 NKLIATRTDHEDDR-----VKTKAFQISMAKPRNSAESGPIYAFENLACEAPPS 908  
 DB 598 RORTLOSKSPADPOVKCVSVRSTALIVSWRPPP--ETHNGALVGY----- 642  
 QY 909 AAHFRYQLEGROYDNTVPFNEDD-----PKSWTEDLIAMPKPMERACIYKVK 959  
 DB 643 SVRYR-----PLGSEDPKEVNGIPPTTQIILEAKETQYRITTVANT 688  
 QY 960 IYGP-----LEVNVSRNMGTHRRIVGKLGIIRDVNRSTRDRDOPNVSAACLEFKCSGMLY 1015  
 DB 689 EYGPRESSPVVYRT-----DEDPVSAAPPRKVE----- 716  
 QY 1016 DDDRVVRLTVKYL-----PGSGRRASVYNMLHEYLVLNHLPLAVNNDTSFYTMLAPL 1067  
 DB 717 -AELALNATIRVLMRSPAPRGHQRIGYQVHYVMEGAEARGPRIKD-----VMLADA 770  
 QY 1068 DPLGHNHYGITYVDDPTAKETALGRCEFGTSDGSSRIKMSNVGVALTFNCVEROVGRO 1127  
 DB 771 QEM-----VIJNLOPETAIVSTIVA-AVTMKDGA-----RSKPKVYVYKGAIV--LGRP 815

QY 1128 SAFQYLOSTPAOS-----PAAGTVQGRVPSRRQGRASRGQROSGVVASLRF 1175  
 DB 816 TL--SVQOTPESSLARKEPPACTAEDQVUGRLQ-----FGNEDSPRLATLEFP 863

RESULT 10  
 US-08-800-825A-5  
 ; Sequence 5; Application US/0800825A  
 ; Patent No. 5866397

; GENERAL INFORMATION:  
 ; APPLICANT: RODAN, GIDEON A.  
 ; APPLICANT: SCHMIDT, AZRIEL  
 ; APPLICANT: RUTLEDGE, SU JANE  
 ; TITLE OF INVENTION: CDNA ENCODING A NOVEL HUMAN PROTEIN  
 ; NUMBER OF INVENTIONS: TYROSINE PHOSPHATASE  
 ; NUMBER OF SEQUENCES: 7  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: J. MARK HAND - MERCK & CO., INC.  
 ; STREET: 126 E. LINCOLN AVE., P.O. BOX 2000  
 ; CITY: RAHWAY  
 ; STATE: NEW JERSEY

; COUNTRY: USA  
 ; ZIP: 07065-0900  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent In Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/800,825A  
 ; FILING DATE: 14-FEB-1997  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: HAND, J. MARK  
 ; REGISTRATION NUMBER: 36,545  
 ; REFERENCE/DOCKET NUMBER: 18992DA  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 732-594-3905  
 ; TELEFAX: 732-594-4720

; INFORMATION FOR SEQ ID NO: 5:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1911 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-08-800-825A-5

Query Match 2.3%; Score 144; DB 2; Length 1911;  
 Best Local Similarity 19.5%; Pred. No. 0.0019;  
 Matches 186; Conservative 122; Mismatches 362; Indels 284; Gaps 46;

QY 295 KAATIKAEFVAETPYVWNPETARRAGOSVSLCKATGKPRDKTFWYHNDLLDPSL 354  
 DB 121 KLTVLREQLPSGFPNIDMGPKLVKERTATMLCAASGNPDE-TWTFDPLVPDPSA 179  
 QY 355 YKHESK-----LVLRKIQHQAGEYFCKAOSDAGAVSKVAQLIVTASDETPCNPDES 408  
 DB 180 SNGRIKQRLSGALQIESSEETDOGKECVATNSAGVRYSSANLYRVYR-----RVAPRF 234  
 QY 409 YLIRLPHDCFQNA-TNSFYDYGR-CP-VKTCAGQDNGICRDVAVONCCGISTEEREI 465  
 DB 235 SILPMSHIMPGANNITICVAVGSPMPYKMMQGAED-----LTPEDDM 278  
 QY 466 QCSGYTLP-TKYAKECSCQCRTESTRSYVGRVSAADNGEPMRFGHYVGNRSVMTGY-K 523  
 DB 279 PVGRNVLELTVDKDSANHPCVAMSSL--GVIEAV-----AQITVKSLEPK 321  
 QY 524 GFTFLHVPQDTERLVLTIVYDRLQKFNVTTKVLPEKKGSAAVHEIKMLRKEPTILEAME 583  
 DB 322 APGTMTVENTATSTITTWDSGNPDVSVYVIEYKSKQDGPYQK-----EDIT 371



```

01 584 TIIIPDGEVDEDMAELEI-----PSSRYRONGEPYIGKVAASYTFIDPRN 631
02 372 TTRYSIG---GLSPSSEYIIVSAVNSIGGCPSESVYTRTGGDAPAR-----PPRN 420
03 632 ISTATAOTDINFINDEGDTPPLRTYGMFVSDFDEVSTSEPLNAGKVKVILDSTGYKMP 691
04 421 VQARLSTMTIYOMEE---PVEPNKILR-GYVYITME-----PE 457
05 692 HISTVKLMSLNPDTGLWEEBGEKFEENORRNKRREDRTFLVGNLEIRERLLNDVESRR 751
06 458 H-----PGWN-----QKHVDDSLTTVGSL-----LEDET 484
07 752 CFVKVRAVRSERFLP-SQI-----QGVISYITNL--EPTTGFLSNPRAG--RPSVY- 800
08 485 YTVRVLATSYGDBGPLSPIDVKTQGGVPCGPMMLRAEASE-TSITLSMPSROEIIK 543
09 801 -----TGNACYPAPACDDQSPDAYSAYVLASLAGEELQAVESSPKFNPAIGVPOPYL 854
10 544 YELFREDDHGREV-----GRTEPPTTSYVEEDLKPTVEYAFRLAAR-SQGGIGATPPV 597
11 855 NKLNTRRTDHDHDP-----YKTAFOISMAKPPRNSAEBSNGEYIAFENLRACEAPPS 908
12 598 RQRTLOSFPASAPDQVCKVSVRSTAILVSWRPPP---ETHNGALVGY-----642
13 909 AAHFERYQIEGDRDYNTVPENED-----PMSMTEDDYLAAMPKPEFRACYIKYK 959
14 643 SVRYR-----PLGSEDPREKEVNGIPTPTTOLLLEALEKMTQYHITTYAHT 688
15 960 IVGP---LEVNVASRNMGTHARTVGYKLYGIRDVSTRDRDQPNVSAACLEKSGMLY 1011
16 689 EYGPDPESSPVYVRT-----DEDVSAAPRKY-----716
17 1016 DQDRDLRLVYI-----POSSCRASVNPMLHELYVNLPLAVANNDTSEXTMLAPL 106
18 717 -AEALNAAIAIRLWKRSPAGRHQOIKGYOYHVYRMGAARGPRIKD-----VMLADA 770
19 1068 DPLGNHYGIVYVTPDDPTAKEILAGRCFDGTSGRSRIKMSNVGVALTFCNVERQVGRQ 1122
20 771 QEM-----VITNQPTATISIVYA-AIYMKKGGA---RSKRKYVYVTGAV---LGR 815
21 1128 SAFVLOYSTPAOS-----PAAGTVQGRVPSRRQORASRGQOROSGVASLRRP 1175
22 816 TL-SVOOTPEGSLIARWEPAGTAEADQVGLGRLO---FGREDSTPLATLEPP 863
23
24 RESULT 11
25 09-158-657-5
26 Sequence 5, Application us/09158657
27 Patent No. 6214564
28 GENERAL INFORMATION:
29 APPLICANT: RODAN, GIDEON A.
30 APPLICANT: SCHMIDT, ARIEL.
31 APPLICANT: RUTLEDGE, SU JANE
32 TITLE OF INVENTION: CDNA ENCODING A NOVEL HUMAN PROTEIN
33 TITLE OF INVENTION: TYROSINE PHOSPHATASE
34 NUMBER OF SEQUENCES: 7
35 CORRESPONDENCE ADDRESS:
36 ADDRESSEE: J. MARK HAND - MERCK & CO., INC.
37 STREET: 126 E. LINCOLN AVE., P.O. BOX 2000
38 CITY: RAHWAY
39 STATE: NEW JERSEY
40 COUNTRY: USA
41 ZIP: 07065-0900
42 COMPUTER READABLE FORM:
43 MEDIUM TYPE: Floppy disk
44 COMPUTER: IBM PC compatible
45 OPERATING SYSTEM: PC-DOS/MS-DOS
46 SOFTWARE: PatentIn Release #1.0, Version #1.30
47 CURRENT APPLICATION DATA:
48 APPLICATION NUMBER: US/09/158.657
49 FILING DATE:
50 CLASSIFICATION:

```

```

1 PROR APPLICATION DATE:
2 APPLICATION NUMBER: 08/800,825
3 FILING DATE: 14-FEB-1997
4 ATTORNEY/AGENT INFORMATION:
5 NAME: HAND, T. MARK
6 REGISTRATION NUMBER: 36,545
7 REFERENCE/DOCKET NUMBER: 189920DA
8 TELECOMMUNICATION INFORMATION:
9 TELEPHONE: 732-594-3905
10 TELEFAX: 732-594-4720
11 INFORMATION FOR SEQ ID NO: 5:
12 SEQUENCE CHARACTERISTICS:
13 LENGTH: 1911 amino acids
14 type: amino acid
15 STRANDEDNESS: single
16 TOPOLOGY: linear
17 MOLECULE TYPE: protein
18 OS-09-158-657-5

```

Query Match	2.38;	Score 144;	DB 4;	Length 1911;
Best Local Similarity	19.58;	Pred. No. 0.0019;		
Matches 186;	Conservative 122;	Mismatches 362;	Indels 284;	Gaps 46

QY	295	KAATIKAEFNAAELEPYVWNPBETKARAGOSVSLCCAKGKPRPKYFENYHNDITLDSL	354
Db	121	KLTVLREDOQLPSGPPNIDMGPOLKLVYERTTATMLCAASGNPDPE-ITMFKDFLIVDPSA	179
QY	355	YKHESK-----LVLRLOOHAGEFCEKAQSDAGAVKSKVQILVTASDETRCPMBES	408
Db	180	SNGRIKOLRSCALQIESSEETDQKTECAVTNAGVYSSPNLVRYR-----RVARF	234
QY	409	YLIRLPHDCEONA-TNSFYVDVGR-CP-VKTCAGQODNOIRCRDAVQMCISGKTEEREI	465
Db	235	SLEPMSEHMPGCVNVTTCVAGSPBRYKMMQGAID-----LTPEDDM	278
QY	466	QCSGYTLF-TRVAKECSCQCTETRSIVGRVNSADNGEPMRGHAYMNSVSTWG-K	523
Db	279	PVGNNVLELTDVKDSANVHCBVMASSL-GVLEAV-----ADITVKSLPK	321
QY	524	GTFELHVPQDTERVLTFVNRLOKFEVTTTFLVLPFNKKSQAVFHEIKMLRKEPTILEAME	583
Db	322	ARGPMYTEMATNSITITTMDSGNPDPRSYVLEIKKSDQGYOLK-----EDIT	371
QY	584	TNIIPLGEVGEDPMABLET-----PSRFYRQNGEPYIGKVKASVTEFLDRN	631
Db	372	TTRRYSIG--GLSPNSEYEIIVASVNSICGGPSESVYRTGEQAPAR-----PPRN	420
QY	632	ISTAMTAQTDLNEINDGDFPLRTGMSVVDREBVTSEPLNAGVKVXHLDSIQVKMPE	691
Db	421	VOARML\$TTIYQWEE---PVEPNGLR-GYRIVYIME-----PE	457
QY	692	HISTVKLMSLNPDTGLAMEEGDEKFEFQNRKKNKEDTFLVGNLEIRERRLFMDVBPESRR	751
Db	458	H-----PYGNM-----QKINVDLSLTTVGSLL-----LEDET	484
QY	752	CFVQVYRATRSERLP-SEQI-----QGVAVSYNL--ERTGELSLNPRAMG--RFDYSI-	800
Db	485	YTVRBLAFTSYGDGPLSDIPLOVKTQOGVPGQPMNLAAEARSE-T\$TSLTWSPPROESIIK	543
QY	801	-----TGPNCACPAFCDDOSPDAVSAYVLASLAGEELQAVESSKPEPNMIGVPOYL	854
Db	544	YELLFREDDHREY-----GRTFDPTTSYVEDLKNTETAYARLARL-SPQSLGAFITYV	597
QY	855	NKLNTRTDEHPR-----VKKTAQI\$MAKPRPNSAEESNCPIYAFENLRACEAPPS	908
Db	598	RORTLO\$KSPAPQDVKVCYVSR\$TALIVSRPPP---ETHNGALVGY-----	642
QY	909	AAHFRFOIBEDRDVNTYVFNED-----PMSWEDVLAMPMPMEBRACIYKVK	959
Db	643	SVYR-----PLGSDDEPPEKVCNGIPPTTQIILLEALEKTOYRITTVYAH	688
QY	960	IVGP-----LEVNV\$R\$M\$G\$THRRYQKLGIDV\$T\$DRDQ\$P\$V\$A\$C\$LFK\$C\$M\$Y	1015

Page 10

```

OY      295 KAATIKAEFVAETPYWVNPEPTKARRAQSIVSLCCCKATGPRPRCKEYVINTDILPSSL 354
          | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      121 KLTVLREDQQLPSPGNIDMPCQLVVERTRTATMTCASGNDDP-ITWEDFLPVPSPA 179
          | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY      355 YKHEK-----LVLRKLOOHQAGEFYCKASQDAGAVKSKVQQLITVASDETPCNPVES 408
          | : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

SIGNED: \_\_\_\_\_  
CITY: New York  
STATE: New York

COUNTRY: USA  
ZIP: 10036  
COMPUTER READABLE FORM:  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/313,288B  
FILING DATE: January 5, 1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P.  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 40028-A-PCT-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 278-0400  
TELEFAX: (212) 391-0526  
TELEX:  
INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1172 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-313-288B-19

Query Match 2.3%; Score 143; DB 1; Length 1172;  
Best Local Similarity 20.8%; Pred. No. 0.001;  
Matches 105; Conservative 56; Mismatches 189; Indels 154; Gaps 29;

QY 86 CARPLRLAET-----TMTFPGSTGQVYVHGSRERFCLNREORPQ 128  
DB 360 CASPFVEGECPCSLHSVDEEGSPWAEWTQCSVT-----CGSGTQQRGR 406  
QY 129 NCSNVTAFILCP-----PGSLRDRTERI-----MSPMSPKSCSAAGOTGVQPTTRICL 178  
DB 407 SCQVTSNTCLPSTIOTRACSLSKCDTRIRQDGMSPMSWSSCSVTGCGVGNTRIRICLN 465  
QY 179 AEMVSL-----CSEASEEGOHCMQDC-----TACDLTCPMG--GVNADCD-- 217  
DB 466 SPVPMGSKNKGSGRETKAGGAPCPIDGRMSPMSACTVTCAGGIRRETRCNPE 525  
QY 218 -----AC-----NC--QDFMLHGAVALP--GGAPASGAALYLLTKPKLLTQDSD 259  
DB 526 PQYGGKACVGVQEROMCNKRSCPDGCLSNPCFPAQCS-----SFPD 569  
QY 260 GFRIRPGICP-----DGSLIKITKVKFAPYVLTMPKTSIKATIKAFVRAETP--YMY 312  
DB 570 GSWSC-GCPVPGFLGNGHCEDLDECALVPDICSTSKVPKCVNTOPGFHCLPCPPRYRG 628  
QY 313 MNPETKARAGOSVSLCKATGKPRDKYFYVYHNDTLDPSSLYKHESKLVLRKLQHQAG 372  
DB 629 NQPVGVGLEAKTEKQVEPE-NPCKDK-----THN-----CHKNAECIYLGHFEDPM-- 675  
QY 373 EYFCKAGSDAGAVSKVAQLIVTASDETPCNPVPESEYILRLPHDFCONATNSFYDV--GR 431  
DB 676 -YKCECQTG-----YAGDGLICG--EDSDLDGPNLILVQATNATATYICIDN 719  
QY 432 CPVATCAQOQDNGIRCRDAVONCC-----GISKTEEREIQCSYTLPTVAKSCS--- 481  
DB 720 CPHLPNSQDEPFD--KQIGIDACDDDDNDGV--TDEKD--NCQLLFNPROADYDKDEVG 773  
QY 482 --CORCTETRSIVRGVSAADNGE 503  
DB 774 DRONCPRVHN--PQIDFTDNNGE 795

RESULT 14  
US-08-469-486-52  
; Sequence 52, Application US/08469486

Patent No. 5739281  
GENERAL INFORMATION:  
APPLICANT: Thoenes, Hans Christian  
APPLICANT: Holte, Thor Las  
APPLICANT: Etzerodt, Michael  
TITLE OF INVENTION: Improved method for the refolding of  
NUMBER OF SEQUENCES: 58  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/469,486  
FILING DATE:  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/192,060  
FILING DATE: February 4, 1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Paul T. Clark  
REGISTRATION NUMBER: 30,162  
REFERENCE/DOCKET NUMBER: 06363/002001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617 542 5070  
TELEFAX: 617 542 8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 52:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 454 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-469-486-52

Query Match 2.3%; Score 142.5; DB 1; Length 454;  
Best Local Similarity 20.1%; Pred. No. 0.012;  
Matches 153; Conservative 73; Mismatches 251; Indels 283; Gaps 39;

QY 132 NYTVAFILCPG-----SLRDRTERIWSFSPWSPKSCSAAGOTGVQTRTRICLAEM----- 181  
DB 2699 NY-----FACPSRCIPMSWTCKEDCEHGEDETHCNKPCSEAFECQNHRCISKQWLCDG 2755  
QY 182 VSLCSEASEEGOHCMQDCQCTACDLTCPMGOVNADCDACMDFMLHGAVALP--GGAPASGA 241  
DB 2756 SDGCGSDDEAHGKTCGSSFSFGPTHY-----CVPERWLDGDKCAQADGADSEIA 2809  
QY 242 AIYLLTKT-----PK-LTQTPSDGFRIRPGICPDGKSLIKITKVFAPIV 286  
DB 2810 AGCLVNSTCDREFMCGQRQCIKHFVCDHNRD-----CADGSD----- 2848  
QY 287 LTMPTSLIKATIKAFVRAETPYVWNPETKARAGOSVSLCKATGKPRDKYFYVHN 346  
DB 2849 -----ESPECEYPTGSPSEFRC--ANGR----- 2869  
QY 347 DTLDPSSLYKHESKLVLRKLQHQAGEYFCKAQSAGAVSKVAQLIVTASDETPCNP-- 404  
DB 2870 -----CLSSRQWECDEGENDCHDQ-----SDEAKNPHC 2897  
QY 405 -VPESEYILRLPHDFCONATNSFYDVGRCPVKT--CAGQODNGIRCRDAV--ONCGISKY 460  
DB 2898 TSPE-----HKC--NMSQFLSSGRCVAEALLCNGQDD-----CGDSDBERGCHINEC 2944

Tue Oct 9 12:15:58 2001

us-09-609-383-2.ra1

Page 12

QY 461 EEREIO-CSGYTLPTKVAKECSQRCCTETRSIVRGVSAADNGEPMRGHYMGNSRVSM 519  
DB 2945 LSRKLGSCSDCEDLKIGKRCRC-----RPGFRLKDDGFRFC-----2980  
QY 520 TGYKGFTHLVPODTERLVLTFVDRLQKFNVTNTKVLFPNKKGSAAVHEIKMLRKEPITL 579  
DB 2981 -----ADVDECSTTF-PCSQRCINT-----HGS-----YKCLC-----3007  
QY 580 EAMETNIIPLEGVYGDPMALIEIPRSFYRONGEPIYKVASVTFIDPRNISTATAQ 639  
DB 3008 -----VEGYAP-----RGDPHSC-----AVTDEEPFLI-----FANR 3036  
QY 640 TDLNFINDEGTEPLRTYGM-----FSVDFRDEVT-----SEPLNAGV-----KVHLDSTQVKMP 690  
DB 3037 YLLKRLNIDGSNTYLLKQGLNNAVALDFYREDMITWDTVTOGSMIRMHNGSNVOYL 3096  
QY 691 EHISTVKLMSINPDTGL-----WEEGDFFKFNORRNRKREDRTFLVGNLEIRER 740  
DB 3097 HRTGL-----SNPD-GLAVDMVGWGNLYWCOKGRDTEVSLNGAY-RTVLVSS-GLREPR 3148  
DB 741 LFNLDVPSRRCFYKVAAYRSEERFLPSEOIOGVIVINLEPRTGLSNPRAMGRFDSVI 800  
DB 3149 ALVDVONG-----YLYWDMGDHSLIGRIGMDSRSYI-----VDTKI 3188  
QY 801 TGPNGACVPACDD-QSPDAYSAYV-LASLAGEELQAVES 838  
DB 3189 TWPNGULTDYTERLYMADAREDIIEFASLDGSRNHVLS 3228  
RESULT 15  
US-08-469-658-52  
Sequence 52, Application US/08469658  
Patent No. 5917018  
GENERAL INFORMATION:  
APPLICANT: Th egeresen, Hans Christian  
APPLICANT: Holte, Thor Ias  
APPLICANT: Ezerodt, Michael  
TITLE OF INVENTION: IMPROVED METHOD FOR THE REFOOLDING OF  
NUMBER OF SEQUENCES: 58  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Fish & Richardson P.C.  
STREET: 225 Franklin Street  
City: Boston  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version  
SOFTWARE: #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/469,658  
FILING DATE: June 5, 1995  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/192,060  
FILING DATE: February 4, 1994  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Paul T. Clark  
REGISTRATION NUMBER: 30,162  
REFERENCE/DOCKET NUMBER: 06363/002002  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617 542 5070  
TELEFAX: 617 542 8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 52:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4544 amino acids

TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-469-658-52  
Query Match 2.3%; Score 142.5; DB 2; Length 4544;  
Best Local Similarity 20.1%; Pred. No. 0.012;  
Matches 153; Conservative 73; Mismatches 251; Indels 283; Gaps 39;  
QY 132 NYTVRFCLPPG-----SLRDRERIMSPWSPKSCSAACGQTGVOTRRIRICLAEM-----181  
DB 2699 NY-----FACPSGRICIPMSWTCDEKEDCEHEDETHCKKFCSEAFQCCQNRHCKQMLCDG 2755  
QY 182 VSLCSEASEGQHCQMODCTACDLTGPMGVNADCDACACODFMILGAVSLPGAPASGA 241  
DB 2756 SDDCGDSDDEAHCHEKTCGPPSSFCPTHV-----CYPERNLCDGDCADGADDESIA 2809  
QY 242 ALYLLTKT-----PK-LLTQTSDSGRPRIPGLCPDGKSLIKTKVFAPIV 286  
DB 2810 AGCLYNSTCDDBREMCQNRQCIRKHFVCDHDDRD-----CADGSD-----2848  
QY 287 LTMKTSLSKAATIAEYVRAETPYMWNMPETKARBAQSVSLCKATGKRPDPKRYWYN 346  
DB 2849 -----ESPECEYPTGCPSEFRC-----ANGR-----2869  
QY 347 DTLDPSTYKHEKSLVLRKLOHQAGYFFCAQSDAGAVSKAQLIVTASDETPCNP--404  
DB 2870 -----CLASSRQWECDEGENDCHDQ-----SDRAPKPNPC 2897  
QY 405 -VPESTYLRPLPDGFRONNPNFYYDVGRCPYKT--CAGQDNGIRCDAY-QNCCGISKT 460  
DB 2898 TSPE-----HKC--NASSQPLCSSGRCAVALLCNGQDD-----CGDSSDERGCHINEC 2944  
QY 461 EEREIO-CSGYTLPTKVAKECSQRCCTETRSIVRGVSAADNGEPMRGHYMGNSRVSM 519  
DB 2945 LSRKLGSCSDCEDLKIGKRCRC-----RPGFRLKDDGFRFC-----2980  
QY 520 TGYKGFTHLVPODTERLVLTFVDRLQKFNVTNTKVLFPNKKGSAAVHEIKMLRKEPITL 579  
DB 2981 -----ADVDECSTTF-PCSQRCINT-----HGS-----YKCLC-----3007  
QY 580 EAMETNIIPLEGVYGDPMALIEIPRSFYRONGEPIYKVASVTFIDPRNISTATAQ 639  
DB 3008 -----VEGYAP-----RGDPHSC-----AVTDEEPFLI-----FANR 3036  
QY 640 TDLNFINDEGTEPLRTYGM-----FSVDFRDEVT-----SEPLNAGV-----KVHLDSTQVKMP 690  
DB 3037 YLLKRLNIDGSNTYLLKQGLNNAVALDFYREDMITWDTVTOGSMIRMHNGSNVOYL 3096  
QY 691 EHISTVKLMSINPDTGL-----WEEGDFFKFNORRNRKREDRTFLVGNLEIRER 740  
DB 3097 HRTGL-----SNPD-GLAVDMVGWGNLYWCOKGRDTEVSLNGAY-RTVLVSS-GLREPR 3148  
DB 741 LFNLDVPSRRCFYKVAAYRSEERFLPSEOIOGVIVINLEPRTGLSNPRAMGRFDSVI 800  
DB 3149 ALVDVONG-----YLYWDMGDHSLIGRIGMDSRSYI-----VDTKI 3188  
DB 3189 TWPNGULTDYTERLYMADAREDIIEFASLDGSRNHVLS 3228

Search completed: October 9, 2001, 11:50:10  
Job time: 265 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 9, 2001, 11:46:55 ; Search time 24.75 Seconds  
(without alignments)  
3644.067 Million cell updates/sec

Title: US-09-609-383-2

Perfect score: 6317  
Sequence: 1 MVTGKAWFSEFLVLEVTSVL.....QSGVVASLRFPYVQAQPLIN 1184

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: PIR-68:\*  
2: PIR1:\*  
3: PIR2:\*  
4: PIR3:\*  
5: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6317	100.0	1184	2 T09484	cartilage intermed
2	169	2.7	1170	2 A40558	thrombospondin 1 p
3	161	2.5	1694	2 S50065	stathmin-like protein
4	157	2.5	1170	1 TSH0P1	thrombospondin 1 p
5	155.5	2.5	7962	2 I38346	elastic titin - hu
6	152	2.4	6658	2 T13931	projectin - fruit
7	148.5	2.4	1125	1 JH0771	protein-tyrosine k
8	147.5	2.3	788	2 T25061	hypothetical prote
9	147	2.3	1444	2 T18856	angiogenesis inh
10	146	2.3	1239	1 A32579	neuroglian - fruit
11	145	2.3	1906	1 S68235	myosin-11ght-chain
12	144.5	2.3	1176	2 JN0583	myosin-11ght-chain
13	143.5	2.3	1034	2 JCS598	mucin - rat
14	143	2.3	1172	1 TSH0P2	thrombospondin 2 p
15	143	2.3	1360	2 T33922	hypothetical prote
16	142.5	2.3	4544	1 S02392	alpha-2-macroglobu
17	142	2.2	1074	1 JCS928	semaphorin F precu
18	142	2.2	3570	2 T45025	mucin MUC5B, trach
19	141.5	2.2	477	2 S53362	mucin 5AC (clone J
20	141	2.2	1256	2 T03096	CDO protein - rat
21	141	2.2	6642	2 T29757	protein UNC-89 - C
22	140.5	2.2	3375	2 T19821	hypothetical prote
23	140	2.2	1122	2 I54237	protein-tyrosine k
24	139.5	2.2	1056	2 A53767	mucin MUC5B, trach
25	139.5	2.2	4391	2 A38096	pelican precursor
26	139	2.2	984	2 T00326	hypothetical prote
27	139	2.2	1522	2 T00028	brain-specific ang
28	139	2.2	2222	2 T13924	sdh protein - fru
29	139	2.2	4162	2 T42633	connectin/titin -

30	138.5	2.2	1123	1 JN0712	protein-tyrosine k
31	138	2.2	1599	2 T16210	hypothetical prote
32	136	2.2	2167	2 T34395	hypothetical prote
33	136	2.2	2907	2 A57278	fibroblast-2 precu
34	135	2.1	1147	2 A59307	myosin-11ght-chain
35	135	2.1	2918	2 A54105	fibroblast-2 precu
36	134.5	2.1	1450	2 A44027	165K myofibrillar
37	134	2.1	1651	2 T14160	transmembrane rece
38	133.5	2.1	1059	2 T22545	hypothetical prote
39	133	2.1	1894	2 C54689	hypothetical prote
40	132.5	2.1	610	2 T16761	protein-tyrosine-p
41	132.5	2.1	860	2 T16892	hypothetical prote
42	132.5	2.1	4545	1 S25111	alpha-2-macroglobu
43	132	2.1	1178	1 A39804	thrombospondin pre
44	131.5	2.1	1742	2 S24600	projectin - fruit
45	131.5	2.1	3707	2 S18252	heparan sulfate pr

ALIGNMENTS

RESULT 1	
T09484	cartilage intermediate layer protein precursor - human
C:Species: Homo sapiens (man)	
C>Date: 16-Jul-1999	#sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
C:Accession: T09484	
R:Lorenzoni, P.; Neame, P.; Sommarin, Y.; Heinigard, D.	
J. Biol. Chem. 273, 23469-23475, 1998	
A:Title: Cloning and deduced amino acid sequence of a novel cartilage protein (CILP)	
A:Reference number: Z16689; MUID:98389785	
A:Accession: T09484	
A:Status: preliminary; translated from GB/EMBL/DBJ	
A:Molecule type: mRNA	
A:Residues: 1-1184 <LOR>	
A:Cross-references: EMBL:AF035408; NID:G3513502; PIDN:AAC33838.1; PID:G3513503	
A:Experimental source: tissue type articular cartilage	
C:Genetics:	
A:Note: CILP	
F:1-21/Domain: signal sequence #status predicted <SIG>	
F:22-1184/Product: cartilage intermediate layer protein #status predicted <MAT>	
Query Match	100.0%; Score 6317; DB 2; Length 1184;
Best Local Similarity	100.0%; Pred. No. 0;
Matches 1184; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
OY 1	MVTGKAWFSEFLVLEVTSVLQRTMLTOSVRRVQGGKKNPSIFAKPADTLESFGEMTTLF 60
DB 1	MVTGKAWFSEFLVLEVTSVLQRTMLTOSVRRVQGGKKNPSIFAKPADTLESFGEMTTLF 60
OY 61	NIDYGGKGDYERLDAIRYYGDRVCARPLRLAARTTMTTPAGSTGVVHGSFRCGFNCL 120
DB 61	NIDYGGKGDYERLDAIRYYGDRVCARPLRLAARTTMTTPAGSTGVVHGSFRCGFNCL 120
OY 121	NREQRRPGNCSWYTRFLCPGSLRDRTERTPSPMSKCAAGCGVQVQRTTRICLAE 180
DB 121	NREQRRPGNCSWYTRFLCPGSLRDRTERTPSPMSKCAAGCGVQVQRTTRICLAE 180
OY 181	NWSLSEASEEGOHMGODCTACDLTCFPGVGNACDCMCDPMHGAVSILPGAPASG 240
DB 181	NWSLSEASEEGOHMGODCTACDLTCFPGVGNACDCMCDPMHGAVSILPGAPASG 240
OY 241	AAIYLLTTPKLLTQTDSDGRFRIPGLCPDGKSLITKRVKAPIVLTPMKTSLKAATIK 300
DB 241	AAIYLLTTPKLLTQTDSDGRFRIPGLCPDGKSLITKRVKAPIVLTPMKTSLKAATIK 300
OY 301	AEFVAETPYVMNPETARRAGOSVSLCCRTGKPRDPKRYWYHNDLLDPSLTKKHESK 360
DB 301	AEFVAETPYVMNPETARRAGOSVSLCCRTGKPRDPKRYWYHNDLLDPSLTKKHESK 360
OY 361	IVLRKLOOHQAGEYFCCKAOSDGAVKSKVAQLITVASDETPCNPPESTLIRLPDCEFN 420
DB 361	IVLRKLOOHQAGEYFCCKAOSDGAVKSKVAQLITVASDETPCNPPESTLIRLPDCEFN 420

Db 361 LVKRLQHQAGEYFCAQSAQSAQAVKVAQLIVTASDEPCNPVSESYLLRLPHDFQN 420  
 QY 421 ATNSFYDVGRCPYKTCAGQDNGIRCDAYONCGISKTREEROCGYTLPTKVAEC 480  
 Db 421 ATNSFYDVGRCPYKTCAGQDNGIRCDAYONCGISKTREEROCGYTLPTKVAEC 480  
 QY 481 SCQCTEYRSIVRGVSAADNGEPMRGHVMGNSRVSMGYSKGTFTLHVPODTERLVLT 540  
 Db 481 SCQCTEYRSIVRGVSAADNGEPMRGHVMGNSRVSMGYSKGTFTLHVPODTERLVLT 540  
 QY 541 FVDRLOKFNVTTKVLPENKKSAAVFHEIKMLRKREPTLEAMENIILPLGVVEDPMAE 600  
 Db 541 FVDRLOKFNVTTKVLPENKKSAAVFHEIKMLRKREPTLEAMENIILPLGVVEDPMAE 600  
 QY 601 LEIPSRSEYRONGEPIYIGKVAASYFLDPRNIISTATAOTDLNFINDEGDTFPLRTYGMF 660  
 Db 601 LEIPSRSEYRONGEPIYIGKVAASYFLDPRNIISTATAOTDLNFINDEGDTFPLRTYGMF 660  
 QY 661 SVDRDEYTSPLNAGVKVYKHLDSYQKMPHISTVYKLMISLNPDTGLMEEGDKFENOR 720  
 Db 661 SVDRDEYTSPLNAGVKVYKHLDSYQKMPHISTVYKLMISLNPDTGLMEEGDKFENOR 720  
 QY 721 RNKREDFTLVGNLEIERERLFINLDPESRCFYKVAYSERFLPSEIOGVVITSYINL 780  
 Db 721 RNKREDFTLVGNLEIERERLFINLDPESRCFYKVAYSERFLPSEIOGVVITSYINL 780  
 QY 781 EPRITGFLSNPRAMGRPDSVITGPGACVPAFCDDQSPDAYSAVYLASLAGEELQAVESSP 840  
 Db 781 EPRITGFLSNPRAMGRPDSVITGPGACVPAFCDDQSPDAYSAVYLASLAGEELQAVESSP 840  
 QY 841 KFNNAIGVPOPYLNKLNRYRTHEDEPRVKTAFOISMARPRPASAESNPITAFENLR 900  
 Db 841 KFNNAIGVPOPYLNKLNRYRTHEDEPRVKTAFOISMARPRPASAESNPITAFENLR 900  
 QY 901 ACEBAPSAAHFRFYQIEGDRYDNTVPFNEDEPMSTEDYLAMPKPMFEFRACYIVYKI 960  
 Db 901 ACEBAPSAAHFRFYQIEGDRYDNTVPFNEDEPMSTEDYLAMPKPMFEFRACYIVYKI 960  
 QY 961 VGPLLEVAVSRNMGTHRRVYKLYGIRDYSTRDRDQPNVSAACLEFFKCSGMLYDDRV 1020  
 Db 961 VGPLLEVAVSRNMGTHRRVYKLYGIRDYSTRDRDQPNVSAACLEFFKCSGMLYDDRV 1020  
 QY 1021 DRTLVKTIPOGSCRASVNPMLHEYLNVNHLPLAVNNDTSEYTMALPDLPGHNGYITVT 1080  
 Db 1021 DRTLVKTIPOGSCRASVNPMLHEYLNVNHLPLAVNNDTSEYTMALPDLPGHNGYITVT 1080  
 QY 1081 DODPRAKETALGRCFPGDSDGSSRIKSNVGVYALFENCYEROVGROSAFOYLOSTPAQS 1140  
 Db 1081 DODPRAKETALGRCFPGDSDGSSRIKSNVGVYALFENCYEROVGROSAFOYLOSTPAQS 1140  
 QY 1141 PAAGTVQGRVPSRROQASRGQROSGVASLPRRYAQQPLIN 1184  
 Db 1141 PAAGTVQGRVPSRROQASRGQROSGVASLPRRYAQQPLIN 1184  
 RESULT 2  
 A:Accession: A40558  
 C:Species: Mus musculus (house mouse)  
 C:Date: 05-Jun-1992 #sequence\_revision 05-Jun-1992 #extl\_change 20-Aug-1999  
 C:Accession: A40558; A37905; B42587; S68787  
 R:Lawler, J.; Duquette, M.; Petro, P.; Copeland, N.G.; Gilbert, D.J.; Jenkins, N.A.  
 Genomics 11, 587-600, 1991  
 A:Title: Characterization of the murine thrombospondin gene.  
 A:Reference number: A40558; MUID:92128941  
 A:Accession: A40558  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1170 <LAN>  
 A:Cross-references: GB:M62449; GB:M62450; GB:M62451; GB:M62452; GB:M62453; GB:M62454; GB:M62455; GB:M62467; GB:M62468; GB:M62469; GB:M62470; NID:9511867; PIDN:AAA5061  
 R:Bornstein, P.; Alt, D.; Devaiah, S.; Framson, P.; Li, P.  
 J. Biol. Chem. 265, 16691-16698, 1990

A:Title: Characterization of the mouse thrombospondin gene and evaluation of the role  
 A:Reference number: A37905; MUID:90375546  
 A:Accession: A37905  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-490 <BOR>  
 A:Cross-references: GB:J05605; GB:J05606; NID:9201991; PIDN:AAA0431.1; PID:9554390  
 R:Lherity, C.D.; O'Rourke, K.; Wolf, F.W.; Katz, R.; Seldin, M.F.; Dixit, V.M.  
 J. Biol. Chem. 267, 3274-3281, 1992  
 A:Title: Characterization of mouse thrombospondin 2 sequence and expression during ce  
 A:Reference number: A42587; MUID:92147683  
 A:Accession: B42587  
 A:Status: preliminary; not compared with conceptual translation  
 A:Molecule type: mRNA  
 A:Residues: 1-1152, 'P', 1154-1170 <LAN>  
 A:Cross-references: GB:M87276  
 A:Note: sequence extracted from NCBI backbone (NCBI:81501)  
 R:Chen, H.; Aeschlimann, D.; Nowlen, J.; Mosher, D.F.  
 FEBS Lett. 387, 36-41, 1996  
 A:Title: Expression and initial characterization of recombinant mouse thrombospondin  
 A:Reference number: S68787; MUID:96234006  
 A:Accession: S68787  
 A:Molecule type: protein  
 A:Residues: 19-26, 'X', 28-37 <CHE>  
 C:Complex: homotrimer, disulfide linked  
 C:Superfamily: thrombospondin 1; EGF homology: thrombospondin type 1 repeat homology;  
 C:Keywords: calcium binding; glycoprotein; homotrimer  
 F:1-18/Domain: signal sequence #status predicted <Sts>  
 F:19-1170/Product: thrombospondin 1 #status predicted <NAT>  
 F:317-375/Domain: von Willebrand factor type C repeat homology <WMC>  
 F:378-429/Domain: thrombospondin type 1 repeat homology <THR1>  
 F:434-490/Domain: thrombospondin type 1 repeat homology <THR2>  
 F:491-547/Domain: thrombospondin type 1 repeat homology <THR3>  
 F:551-586/Domain: EGF homology <EGF>  
 F:248,360,708,1067/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 2.7%; Score 169; DB 2; Length 1170;  
 Best local similarity 20.6%; Pred. No. 0.0028;  
 Matches 113; Conservative 38; Mismatches 151; Indels 246; Gaps 28;

QY 10 SFLVLEVTSLVGLQMTLQSVRRVQEK-----KNPSFAKPAOTLSPGCGWTT----- 58  
 Db 278 SSMVLELKGRLTIVTLQDSIRKTEENRELVSRLRPLCFHNGVQYKNEEMVTDSC 337  
 QY 59 -----LFNIDYPGKGDEYERLDAIRFYGGDVCARPLLEARTTD 98  
 Db 338 ECHGQNSVITICKRVSCPIMCSNATYPDE-----CPRCWPSSADG 381  
 QY 99 WTPAGSTGVVHSGSPRGEW-----CLNEORPGQNSNYTVRLCPGSLRRTERT- 151  
 Db 382 WSPWSE-----WTSQATGNGIQGRSCDSLNRR--CEGSSVQTRCHIQ 426  
 QY 152 -----WSPWSPWKSASAAGOTGVQTRTICLAEMVSL-----CSEASEEOHOM 196  
 Db 427 ECKRREKQDGMWSHSPWSSCVTCGD-GVITRIRLCSNPSPOMNKPKCEGEARETRKAK 485  
 QY 197 GQDC-----TACDLTC-----PMGQVNA-DC-----DACMCD 223  
 Db 486 KDACPINGMGWPMSPMDICSVTCGGVQRRSRCLNPTQFGKDCVGVTEQVNOYKNKD 545  
 QY 224 FMHAGVSLPGCAPASGAATLYLTTPKLLTQDSGRFRIRPCLDGSILKTKVKFA 283  
 Db 546 CPIDGCLSNPCFAGAR-----CTSTP-----DGSWKC-GACPPGYS----- 580  
 QY 284 PIVLTPKTSLKATATKAETVRAETPYMMNPTKARRAGQSYSLCKKATG--KRPDPK 341  
 Db 581 -----GNGIQ--CKDVDECKEVPDAG 599  
 QY 342 FWHYNDTLDPSSLYKHESKIVLRLQHQAGEYFCAQSDAG----- 383  
 Db 600 F-NHN-----GERHCK-NTPDPVNCILPCPRPRFTGSGPQR 632



F:434-490/Domain: thrombospondin type 1 repeat homology <THR2>  
 F:491-547/Domain: thrombospondin type 1 repeat homology <THR3>  
 F:510-586/Domain: EGF homology <EGF1>  
 F:650-689/Domain: EGF homology <EGF2>  
 F:926-928/Region: cell attachment (R-G-D) motif  
 F:171-232/Disulfide bonds: #status predicted  
 F:248-360,708,1067/binding site: carbohydrate (Asn) (covalent) #status predicted  
 F:270-274/Disulfide bonds: interchain #status predicted  
 F:610/Modified site: erythro-beta-hydroxyasparagine (Asn) #status predicted  
 F:1051/Binding site: carbohydrate (Asn) (covalent) #status absent

Query Match 2.5%; Score 157; DB 1; Length 1170;  
 Best Local Similarity 20.1%; Pred. No. 0.019;

Matches 111; Conservative 37; Mismatches 150; Indels 254; Gaps 27;

QY 10 SFVLVETVSLGKQMTLQSVRYQPKNPSTFAKPADLTLESP-----GEW 56  
 DB 278 SSMVLELRLGLRTVTLQDSIRKYTEENK-----ELANELRRPPLCYHNGVOYRNNEEW 331  
 QY 57 TT-----LEFIDYPGKGDYERLDAIRFYGDVRCARPLRL 92  
 DB 332 TVNSCEHCQNSVTICKKVCSPIMCSNATVDEG-----CCPRCWP 375  
 QY 93 EARTTWTPTAGSTGVVHGSPPREGFW-----CLNREQRQNCNNTVRLCPPLSLRR 146  
 DB 376 DSADDGSPWSE-----WTSCSTSCGNGIQGRSODSLNNR--CEGSSVQT 420  
 QY 147 DTERI-----WSPSPWMSKCSAAGCGTGVOTRTRICLAENVSL-----CSASE 130  
 DB 421 RTCHIDECDRKFKODGWSHSPWSSCSVTGCD-GVITRIRLCNSPSPQNGKRCGEAR 479  
 QY 191 EGOHCMGDC-----TACDLTC-----PMGOVNA-DC-----D 217  
 DB 480 ETACCKKDCAPINGGMPSPMDICSTYCGGQVQKSRCLNNPFPQGGDCYDVTENQ 539  
 QY 218 ACMQDPMHLGAVSLPGADASGAAYLLTKTKPLLTQTDSDGRFRIRPLGCPDGKSLTKI 277  
 DB 540 ICKKQDCPDIGCLSNCFACVK-----CTSP-----DSMKC-GACPPGYS---- 580  
 QY 278 TKKFAPIVLTMTKTSIKATIKAEVRAETPYMWNPETKARRAGSVSLCKATGKPR 337  
 DB 581 -----GNGIOCTDVDECKEY 595  
 QY 338 PDKFFWYHNDTLDPRLSKHESKLVRLKLOHQAGEYFCKAQSADG----- 383  
 DB 596 PDACF-NHN-----GEHRCE-NTDPGNCPLCPRPRTGSQ 628  
 QY 384 ----AVSKVYAQLIVTASDETPCNVPYESTLRLPHDCFONATNSF---YVD-VGRCPVK 435  
 DB 629 PFGQVEHATANKQVC---KPRNPCTDG---THDCNKNACNYLGHYSDEMYRCECK 679  
 QY 436 TCAGQDNGIRG 447  
 DB 680 --PGYAGNGILIC 689

RESULT 5

138346 elastic titin - human (fragment)

C:Species: Homo sapiens (man)

C:Date: 29-May-1998 #sequence\_revision 29-May-1998 #text\_change 21-Jul-2000

C:Accession: I38346

R:Label: S.; Kolmerer, B.

Science 270, 293-296, 1995

A>Title: Titins: giant proteins in charge of muscle ultrastructure and elasticity.

A:Accession: I38346

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues:1-17962 <RES>

A:Cross-references: EMBL:X90569; NID:g1017426; PIDD:CAA62189.1; PID:g1017427

A:Gene: GDB:TTN  
 A:Cross-references: GDB:127867; OMIM:188840  
 A:Map position: 2q31-2q31

Query Match 2.5%; Score 155.5; DB 2; Length 7962;

Best Local Similarity 17.5%; Pred. No. 0.4;

Matches 197; Conservative 139; Mismatches 353; Indels 437; Gaps 51;

QY 17 TSVLGKQMTLQSV-----RRVQPKKNPSTFAKPADLTLESGEWTTL-----FN 61  
 DB 2529 SSDMGNTVCYANAVAGSDECRALVTQEPSPFVKEPEPEPLEVLGKNVTFVIRGTPPEK 2588  
 QY 62 IDYPGKGDYERLDAIRFYGDVRCARPLLEARTDWTMPAGSTGVVHGSPPREGFWCLN 121  
 DB 2589 VNNFRGARELVKGDRCNITYEDTVA---ELELFNIDISQSGEYCVSNNAQAS-CYT 2643  
 QY 122 R-----EQRPQNC--SNVTRFLCPPLSLRDRTERIWSPPWS-PWSK- 160  
 DB 2644 RLFVKEPAFLKRLSDHSVSPGKSIILESTYT-----GTL-----PISVTWKD 2687  
 QY 161 ----CSAAGCGTGVOTRTRICLAENVSLCSSEASBSQHCMDCTACDLTCPMGOVNA 215  
 DB 2688 GFNITTSSEK--NIVTTEKTCILEILN--STRKDAQY-----SCEIENAGR---- 2731  
 QY 216 CDACMCDPMLHGAVSLPGAPASGAAYLLTKTKPLLTQTDSDGRFRIRPLGCPDGKSL 275  
 DB 2732 -DVC-----GALVSTL----- 2741  
 QY 276 KITKVKFAPIVLTMTKTSIKATIKAEVRAETPYMWNPETKARRAGSVSLCKATGK 335  
 DB 2742 -----EPYFVTELEPLAAGVDSVSLQCVAGT 2770  
 QY 336 PRDKFFWYHNDTLDP-----LYKHSKLVRLKLOHQAGEYFCKAQSADGAVSKV 389  
 DB 2771 PE-ITVSWKGDYKRLRTPRYFTFTNNVATLVENKVININDSEYTKAENSIGTASSTK 2829  
 QY 390 AOLI-----VTASDETPCNVP-----ESTYLRIRPHDCFO- 419  
 DB 2830 VFRIQERQLPSPARQLKDIEQTVGLPVTITCLNLSAPLQVWYDGVLLR-DHENLOT 2888  
 QY 420 -----NATN-----SFYYDVGRCPVKTC 437  
 DB 2889 SFVDNVAATLKILQDLSHSQSGSASNPGLTASSARLARARPKKSPFDIKPVSDIVY 2948  
 QY 438 AGQDNGIRCRDAVQCCGS-KTEEREIQCG-YILP-----TKVAKEGCG-Q-R 484  
 DB 2949 AGESAD-FECHVTGAOPMRITWSKDKKEIRPGGNVYITCVGNTPHLRILKVGCGDGYT 3007  
 QY 485 CTEFTRISIVGRVSA-----ADNGEPMRFGHYVMGNSRVSMGTGKGT 525  
 DB 3008 CQATNDVGMCSAQLSVKPPKFKKLEASKYAKAGESQLECKISGSEIIVSMFRND 3067  
 QY 526 FTLHVQDTERLVLTIVDRLQKFNVTTKVLPNK----- 559  
 DB 3068 SELHESMKYV-----MSFINSVALLTINESAEDSGDYICEAHNGVDASCSTAL 3117  
 QY 560 --KSAVFHEIKMLRRKEPTITLEAME-TNIIPLGEVVGEDPMALPIPSRSFYRGGEY 616  
 DB 3118 TVKAPPVF-----TQKSSPVGALGSDVILLOCELSGPPPEVWVKDKOYR-NSKRF 3169  
 QY 617 IGRKVASVTFLEDRNISTATAQTDLNFINDG-DTF-----PLRTYGMFS----- 661  
 DB 3170 KIRSKHEDTNLHLINLEASDVGEYHCKANNVGSJDCSCSVKKEPRPVKKLSQSTSLI 3229  
 QY 662 ----VDRDEYTS-EPLNA-----GKVYHLDSQYKMPBHISTVLMG----- 700  
 DB 3230 GDVAVELRAIVEGFQPISTVWMLKDRGEVIRESENTRISFIDNIATLQSGPEASNSGKYIC 3289  
 QY 701 -LNPDRGLMEBEDGFENQRR--NKREDRTFLVGN----- 733  
 DB 3290 QIKNDAGMRCSAVLTVLEPARITTEPDPATVTTGPFALCEVVTGTPELSAKWFKDGE 3349



```

OY 734 -----LEIRERRLNLDVPE---RRCVVKVARESERELPS- 767
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 3350 LSADSKHHITFINKVASLKIPCAEKSMDGLVSEFKNSGNCIVSYHV--SDILVPPS 3407
OY 768 -----EQLOGVVISYINLEPRGFLSNBRAMGRF---DSVITGPGACVPAFCD----- 813
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 3408 FIRKLKDNALIGASVLEECRSG--SADISYGMFODGNEIVSGPK--COSSFSENVCTLN 3464
OY 814 -----DQSPDAYSAYVLASLAGE---LQAVESPKF--NPALIGV 849
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 3465 LSLPEPDTGITTCVAAVAVAGSDSCSAVLTVQEPSPSEQTPDSVEV 3510

RESULT 6
T13931
projectin - fruit fly (Drosophila melanogaster) (fragment)
C:Species: Drosophila melanogaster
C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 17-Nov-2000
C:Accession: T13931
R:Daley, J.; Southgate, R.; Ayme-Southgate, A.
J. Mol. Biol. 279, 201-210, 1998
A:Title: Structure of the Drosophila projectin protein: isoforms and implication for pro
A:Reference number: Z17815; MUID:98300339
A:Accession: T13931
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-6658 <DAL>
A:Cross-references: EMBL:AF047475; NID:g3337430; PID:g3337431; PIDN:AAC27550.1
C:Genetics:
A:Gene: projectin
A:Cross-references: FlyBase:FBgn0005666
A:Map position: 4
A>Note: Intron positions not resolved (incomplete sequence)
C:Keywords: muscle

```

```

Query Match 2.4%; Score 152; DB 2; Length 6658;
Best Local Similarity 19.6%; Pred. No. 0.54;
Matches 162; Conservative 100; Mismatches 259; Indels 306; Gaps 40;

```

```

OY 234 GAAPASGAI-----YLLTKPKLLTQTD-----SDGRFRIPGLCPDGK---- 272
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 1742 GGASTIGYVERKDPNTGKWKQKALESTPDCARVNDIAGNKKYQFRIMAVNKAAGSKSPS 1801
OY 273 --SILKIKVAFAPIVLMPK---TSLKAATIKAEFVRAETPYWMANPETKARRAGQSS 327
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 1802 EPSDQMTAKDRAP-----PKIDRTNIDITSK-----AGQHTR 1835
OY 328 LCCATGKPRPKYFMYHNDTL--DPSLY-----KHESKLYLRKIQOHAGEYRCQAS 380
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 1836 FDIKVSGER-PATKWLINNKARLENDSDNYNIDMESYRKLTVPISKRHSKYLTKAEN 1894
OY 381 DAGAVKSVADLITVASDETPCNVPESYLRLPHDCFQMATNSEFYDVGK--CPVKTC 438
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 1895 ESGRDEASFEVILVD-----KPGPREGPLRVT-----DVHKRGCKLKNWA 1934
OY 439 GQDQNGICRAVONCCGISTKEEIEIQCSTGLPTKYAKECSQ--KCTETSIYGRKVS 497
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 1935 PLDDGGLPIDHYI-----IEKMDVE--SGHWLPSGRKESFAELNNLEPSHEXKFRVL 1985
OY 498 AAD---NEPMFGRHVMGNRSFVSWTGYKGFTHLVPODTEHLVLFVDRLOKFNNTKV 554
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 1966 AVNTGSESEPLTGEQSVLAKNPFDEPGKPT-----PE-----AYDMQKHVDLWVR 2032
OY 555 LPFNKGSVAVHEIKMLRKKEPT--LEAMETNIIPGE-----VVGEDPMALLET- 603
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 2033 PPIINDGSGPITGV--VEKREKGTAKIKGEITIPICLGECKATVPTLNCGEYEFVK 2090
OY 604 -----PSRSFYRNGEPIYIGV--KASYFLDP--NISTATAAQDILN--INDEGD 650
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 2091 AINAAFGPSPDA-----SKPIITPKRIAPITLDPNTKINIRYFKSGEPIFLDINISGE 2145
OY 651 TEPRLT-----YGMFSVDR 665

```

```

Db 2146 PAPDVTWNNKNSVQTTSPSHIENLPYNTKIINNPERKDTGLYKISAHNFYGGQOQVERQ 2205
OY 666 DEVTSEPLNAG-----KQVHLDSTOYKMP-----EHISTYKLMSLNDPTGLMEBEGDF 714
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 2206 INIITKPGKPGPGLLEVSEHKGDKLKKPKDKDGEPEVSESYLVERKFPDPTGIWLPVG-- 2263
OY 715 KFEQNRNKRREDRTFLVGNLEIRERRLNLDVPESRRCVVKRAYRSEFPLPSEDIOGVV 774
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 2264 -----RSDGE-----YVNDGLVGHQYK--FR 2284
OY 775 ISVINLEPRTGLSNBRAMGRFDSVITGPGACVPAFCDQSPDAYSAVYASLAGEELQ 834
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 2285 VKAVNKE-----GESEPLET-----LGSIIAKDPF 2309
OY 835 AVESSPKENPNAIGVPOPLYKLNRDHDPRVKTAFOISMARPNASAEESNGPI- 893
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 2310 SVPTKRP-----GVPEP-----TDW-----YANKVELAMPFP--ASDGSPIQ 2344
OY 894 -YAFENL-----RACEAPSAHPRFYQIEGDRDYNTVPNPE 931
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 2345 GIVEYKDKYSPLMEKALETNSPTPTATVQGLIEGNEYQFRVALNK 2391

```

```

RESULT 7
JH0771
protein-tyrosine kinase (EC 2.7.1.112), receptor type tyk precursor - mouse
C:Species: Mus musculus (house mouse)
C>Date: 10-Jun-1993 #sequence_revision 10-Jun-1993 #text_change 21-Jul-2000
C:Accession: JH0771
R:Horita, K.; Yagi, T.; Kohmura, N.; Tomooka, Y.; Ikawa, Y.; Aizawa, S.
Biochem. Biophys. Res. Commun. 189, 1747-1753, 1992
A:Title: A novel tyrosine kinase, tyk expressed in murine embryonic stem cells.
A:Reference number: JH0771; MUID:93129253
A:Accession: JH0771
A:Molecule type: mRNA
A:Residues: 1-1125 <HOR>
A:Cross-references: GB:D13738; NID:g220439; PIDN:BA02883.1; PID:g220440
C:Genetics:
A:Gene: tyk
C:Superfamily: protein-tyrosine kinase, receptor type tie; EGF homology; fibronectin
C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phos
F:1-22/Domain: signal sequence #status predicted <SIG>
F:23-1125/Product: protein-tyrosine kinase, receptor type tyk #status predicted <MAT>
F:37-104/Domain: immunoglobulin homology <IM1>
F:133-137/Region: cell attachment (R-G-D) motif
F:212-252/Domain: EGF homology <EG1>
F:256-299/Domain: EGF homology <EG2>
F:303-341/Domain: EGF homology <EG3>
F:365-427/Domain: immunoglobulin homology <IM2>
F:448-526/Domain: fibronectin type III repeat homology <FN3A>
F:541-625/Domain: fibronectin type III repeat homology <FN3B>
F:638-720/Domain: fibronectin type III repeat homology <FN3C>
F:752-773/Domain: transmembrane #status predicted <TM>
F:823-1100/Domain: protein kinase homology <KIN>
F:831-839/Region: protein kinase ATP-binding motif
F:1140,138,400,439,465,550,556,669,691/Binding site: carbohydrate (Asn) (covalent) #st
F:856,873,965/Active site: Lys, Glu, Asp #status predicted

```

```

Query Match 2.4%; Score 148.5; DB 1; Length 1125;
Best Local Similarity 17.7%; Pred. No. 0.069;
Matches 174; Conservative 117; Mismatches 320; Indels 373; Gaps 40;

```

```

OY 10 SFLVEVTSVLGRQTMLTQSVRRVOPGKKNPISIFAKPA--DTLESFG--EWTLTFNIDYFG 67
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 123 SFLPATLFTVTRGDNVNISFKVLIKEDAVIYKNGSLHPLSARGMKYLDILEVHLRHA 182
OY 68 KQDYERLDAIRFYGYDRVCARLREARTD---WTPACSTQGVVNGSPREDFCLNREQ 124
| : : : : : | : : : : : | : : : : : | : : : : : |
Db 183 QPDAGVYSARYIGNLFSAFTLIVRCEAQKKGPDS----- 222
OY 125 RFGQNSNV-----TVRLCPGSLRDPTRTERIWPSPWMSKCSAAG--GQTGVOTRTRIC 177

```

```

Db 223 RPTCTCKNNGVCHBDTGCICPPEFMGRTCEKACEPHTFGRCTKERCSGPEGKASY-FC 281
178 LAENVSLCSEASESGQHMGDDCTACDLTCMGVYMAC-----DACMCQ 222
282 LPDYYG-CSCAT-----GWRGLQCNEMACPSGGYGPCKLRCHCTNEICDRFGGLCS 333
223 DFMHGAVALFEGAPASGAALYLLTKPILKLTQTDSDGRFRIPGLCPDGKSLKITKVKF 282
334 -----QMGGL-----QCEKGRPMPTQIIDLPHIEVNSKF 367
Db 283 APIYLYMPKTSLKAAITKAETVRAETPYMMNPETKARRAGQSVLCCCKATGKPRP--D 339
368 NP-----CKASGMPLEPTSEE 383
340 KYFVYHNDLLDPSLYKHESKLVLRKIQHQAGEYFCKAQSDAGAVSKVQQLVTASDE 399
384 MTLVKPGGYIQLPNDFTYDF-----SALFTVNRV 415
400 TPCNPVPESTYLRPHDCFOATNSFYDVGRCVCKTACAGQDNGIRCAVQNCGSIK 459
416 LP-----PDS-----GVVCSVNTVAGM----- 433
460 TEEREIQSGTLEPTKYAKECSCQCTETSIYGRVRSADNGEPARFGHY--MGNS--- 515
434 -VEKFNISYKVLK-----EPLHAPNVITDGHAFAI 463
516 -RVSMGTGKGFTHVPODFEELVTFVDRLQKQVNTTKVLPENKGSAYHEIKMLRK 574
464 INISSEPFYF--DGPISKKLFPYVQNAKRIEY-----NEFTLNTYL 506
575 EPTLEAMEFTNIIPLGEVVGEDPALEIDPSRFYRONGEPIYGVK---ASVTFLLDP 630
507 EPRDYELCYQLAPGE-----GGEHPFPVRRFTTASIGLPPR 546
631 NISATGAOTDLN-----FINDEGDTFPLRTYGMFSVDPRDEYTSPLNAGKVKVILDS 684
547 GLSLPSSQALNLTWQPIFTNSD-----EFVEVERSLQTS---DQ 588
685 TOVKMPEHISTVKLMSLN-----DTGLWEE-----GDF--KEENORRN 722
589 QNIKVPENLTVSLNSLVPREQYTVARAVNTKAQGEWSEELRAWTLDDLPRQENIKIS 648
723 KREERTFLVGNLEIRERLNLV-----PESRCFVYVRAVRSRFLPSBEOIGVVI 775
649 NITDSTAMVMTIVDYSISITIRYVQGNEDQHDIVKIKATVQY---QLKG--- 701
776 SVINLEPRTGFLSNPRAMGRFDSVITGNGACVPAFCDD-----QSPDAYSAVYLAISAG 830
702 -----LEBETTY-----HVDIFAENNIGSSNPAFSEHLRTILPHSPAS-----ADLGG 743
831 EELQAVSSKFNPNNAICVP--QPYLKLKLYRRTD-----HEDPRVAKTAFOI 876
744 GK-DATSHHWVGNWDASPCLLAFILMLQKRAVQRRMAQAFQVNRREPAVQFNGSL 802
QY 877 SMAKPRNSAEESNGPIYAFENLR 900
Db 803 ALNRKAKNNDPPTIYVLDWMDIK 826

```

RESULT 8  
 T25061  
 hypothetical protein T21B6.3 - *Caenorhabditis elegans*  
 C:Species: *Caenorhabditis elegans*  
 C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 18-Feb-2000  
 C:Accession: T25061  
 R:cottage, A.  
 submitted to the EMBL Data Library, November 1995  
 A:Reference number: Z19975  
 A:Accession: T25061  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-788 <WIL>

A:Cross-references: EMBL:Z68011; PIDN:CAA92014.1; GSPDB:GN00028; CESP:T21B6.3  
 A:Experimental source: clone T21B6  
 C:Genetics:  
 A:Gene: CESP:T21B6.3  
 A:Map position: X  
 A:Introns: 20/1; 47/1; 76/1; 152/1; 735/2; 754/2

Query Match 2.3%; Score 147.5; DB 2; Length 788;  
 Best Local Similarity 27.2%; Pred. No. 0.048;  
 Matches 55; Conservative 15; Mismatches 65; Indels 67; Gaps 11;

```

QY 46 PADLESPEWTTLEFNIDYPGKGDYERLDAIR-----YGGRVCAAPLRLE 93
Db 501 PCQWSEMCWST---CSASGSGQRER---TRCHLGTNRCEKDYESQCSAGPCPEW 554
QY 94 ARTDWTAGST-GQVYHGSFREGFWCL-----NREOR--PGNCSNYTVRFL 138
Db 555 SQWEDMGCCSVTCGQAVVQRT---CLGGVFQDHLCCGPKTEQRACDGGPCS----- 604
QY 139 CPPESLRRDTERIWPSPWSPKSCAAGQTVGQVOTRTRICLAENVSLCSEASESGQHMGQ 198
Db 605 -----LMSFWQEMSTCSASC--SGMKRQKRV--QETDCQGPNESSQFCYGP 649
QY 199 DC-----TACDLTCPMGQ 211
Db 650 PCAEWTEWCEWSSGSSKCGPGQ 671

```

RESULT 9  
 T18856  
 angio genesis inhibitor homolog - *Caenorhabditis elegans*  
 C:Species: *Caenorhabditis elegans*  
 C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 18-Feb-2000  
 C:Accession: T18856; T24653  
 R:McMurray, A.  
 submitted to the EMBL Data Library, July 1995  
 A:Reference number: Z19031  
 A:Accession: T18856  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-1444 <WIL>  
 A:Cross-references: EMBL:Z50004; PIDN:CAA90293.1; GSPDB:GN00028; CESP:C02B4.1  
 A:Experimental source: clone C02B4  
 R:McMurray, A.  
 submitted to the EMBL Data Library, July 1995  
 A:Reference number: T24653  
 A:Accession: T24653  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-1444 <WIL>  
 A:Cross-references: EMBL:Z50006; PIDN:CAA90302.1; GSPDB:GN00028; CESP:C02B4.1  
 A:Experimental source: clone T07C5  
 C:Genetics:  
 A:Gene: CESP:C02B4.1  
 A:Map position: X  
 A:Introns: 25/3; 70/3; 96/3; 139/3; 187/1; 234/2; 282/3; 376/2; 422/2; 478/3; 509/3;

Query Match 2.3%; Score 147; DB 2; Length 1444;  
 Best Local Similarity 28.9%; Pred. No. 0.13; 64; Indels 22; Gaps 8;  
 Matches 44; Conservative 22; Mismatches 64; Indels 22; Gaps 8;

```

QY 86 CARPLRLAART-----TDWTP-AGSTGVYHGSFREGFWCLNREORPGQNCSTYVR-F 137
Db 1297 CFPDPARCLHLDGMSWTMSDWTPCASGSGFQVQNRDS---CSSPFRKGQSGCGLAHQNS 1353
QY 138 LCPPGSLRRDTERIWPSPWSPKSCAAGQTVGQVOTRTRICLAENVSLCSEASESGQHMG 197
Db 1354 LCLDLPACDHESDGEWSAMNEMSGCMGNC--IGTRTVRAVSPV-----SDGQGPFCG 1406
QY 198 QDCTACDLT-CPMGVYVADDCDACQDFMLHG 228

```

Db 1407 R--SSEITECRQSPSTALCSSFITSHLADG 1435

RESULT 10

A32579

neuroglian - fruit fly (*Drosophila melanogaster*)

C:Species: *Drosophila melanogaster*

C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999

C:Accession: A32579

R:Blieber, A.J.; Snow, P.M.; Hortsch, M.; Patel, N.H.; Jacobs, J.R.; Traquina, Z.R.; Schi Cell 59, 447-460, 1989

A:Title: *Drosophila* neuroglian: a member of the immunoglobulin superfamily with extensiv

A:Reference number: A32579; MUID:90030418

A:Accession: A32579

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-1239 <BIE>

A:Cross-references: GB:M8231; NID:g157998; PIDN:AAA28728.1; PID:g157999

A:Note: the authors translated the codon TAT for residue 1234 as Thr and AAA for residue

C:Genetics:

A:Gene: FlyBase:Ng

A:Cross-references: FlyBase:FBgn0002968

C:Superfamily: neural cell adhesion molecule LI; fibronectin type III repeat homology; I

C:Keywords: alternative splicing; cell adhesion; duplication; membrane protein

F:353-412/Domain: immunoglobulin homology <IMM1>

F:446-502/Domain: immunoglobulin homology <IMM2>

F:535-596/Domain: immunoglobulin homology <IMM3>

Query Match 2.38; Score 146; DB 1; Length 1239;

Best Local Similarity 20.38; Pred. No. 0.12;

Matches 149; Conservative 99; Mismatches 277; Indels 208; Gaps 41;

Qy 304 VRAEPYVWNPETKARAGOSVSLCKATGKRPDKYFWHNDLDPSTLYKHSK--L 361

Db 427 VQAEPTTISEAPAAVSTYDGNVITKCRVNGSPKP-LVKWILASWMLTGGRRVNOANDL 485

Qy 362 VLRLQHQAGPEPCAKQSDAGAVSKAQLIVTASDETPCPVPESVLIIRPHCFQNA 421

Db 486 EIQDFTSDAGYTCYCAONKRGELQAD-GSLV--KEHRTIQTQONTEVA---AGOSA 538

Qy 422 T---NSFYDVGRCPCVTKCA-GQ-----QDNGIRCDAVONCGISTEEREICSGYT 471

Db 539 TPCREAHDLLEIFIDMKKQGSIDFPAQPRFVKTND---NSLTIAKT--MELDSGETY 593

Qy 472 LPTKYAKESCORCTETRSIVR-----GRVSAADNGEPRRGHYMGNSRVSMGTG 522

Db 594 C---VARTRLDEATARANLIVQDVPNAPKLTGICQADKAE---IHWEGQGNRSPIILHY 647

Qy 523 KGTFTLHVP-----QDTERLYTFYDRLOKFPVN-TTKVLPRKKGSA--VFHEIKML 571

Db 648 TIOFTSTTPASMDAAVYKRPVNTDSSFYVQMSPMANLYFRVIAFNKIGASPSASDSCT 707

Qy 572 RR-----KEPTLEAMENIIPLEGVEGDPMAELE-----TPSRSF-- 608

Db 708 TOPDVPFKNPDNVVQGGF--PNNLYISWTMPETLEHNAPNHYVSKKRIIPAAWENN 765

Qy 609 ----YRQNG-----EPYIGKVKASYFLDPNRISTATAQTDLNFINDEGDTFPLRTYGM 659

Db 766 NIFDRQNNIIVADQPTFVKYLIKVAINDRGESNVAAEEV---VGXGEDRPIDAPLN 821

Qy 660 FSVDRDVTSE-----PLNKGKVKVHLDSTQVKMPEHISTYKWLMSLNDTGLMEE 711

Db 822 FTM--RQITSTSGYMAWTPVSESVKRGKGYK-----IQTWT-----ENE 861

Qy 712 GDFKEENOR--RNKREDTFIVGNLEIRERRLFNLDVPSRRCFYKAVAYRSERFPSQ 769

Db 862 GE---EGAREIHVKGDTHNALVTQK-----PDSKN-VARILAYNR----- 899

Qy 770 IQGVVIVINLEPRTGFLSNPRAMGRFDSVITGPNACVPAFCDDQSPDAY---SAYVL 825

Db 900 -----FNGCPSPAVIDPDT-----PEGVPSPV---QGLDAYVPLGSSAFWL 935

Qy 826 -----ASLAGELOAVESSPKFNPAIGVPOPLYLNKLYRRTHEDPPKKTAFOI 876

Db 936 HMKKPLYPNGKLTLYKRIYEE-----VKSEYGERREYDPHTDPVTR-----M 980

Qy 877 SNAKRPNS-----AESNGPIYAFE--NLRACEAPPSAAHFRFYQTEGD----R 921

Db 981 KNAGLKPNKSKYRISTATATTKMGEGSEHYIEKTTLKDAVAVAPATPSFWEQLPSDNLAK 1040

Qy 922 YDYNVPFNEDDP 934

Db 1041 FRINWLPSTEGHP 1053

RESULT 11

S68235

myosin-light-chain kinase (EC 2.7.1.117), 210K, nonmuscle - chicken

N:Contains: myosin-light-chain kinase, 108K, smooth muscle; telokin

C:Species: *Gallus gallus* (chicken)

C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999

C:Accession: S68235; A37099; B44389; S28227; S78216; A35093; A25810; S11652

R:Materson, D.M.; Collinge, M.; Lukas, T.J.; Van Eldik, L.J.; Bitukov, K.G.; Stepano FEBS Lett. 373, 217-220, 1995

A:Title: Multiple gene products are produced from a novel protein kinase transcriptio

A:Reference number: S68235; MUID:96033976

A:Accession: S68235

A:Status: nucleic acid sequence not shown

A:Molecule type: mRNA

A:Residues: 1-1906 <MAT>

A:Cross-references: EMBL:X52876; NID:g992992; PIDN:CAA37056.1; PID:g992993

R:Shoemaker, M.O.; Lau, W.; Shattuck, R.L.; Kwiatkowski, A.P.; Matrisian, P.E.; Guerr J. Cell Biol. 111, 1107-1125, 1990

A:Title: Use of DNA sequence and mutant analyses and antisense oligodeoxynucleotides activity.

A:Reference number: A37099; MUID:90361738

A:Accession: A37099

A:Molecule type: mRNA

A:Residues: 649-1906 <SHO>

A:Cross-references: EMBL:X52876

R:Collinge, M.; Matrisian, P.E.; Zimmer, W.E.; Shattuck, R.L.; Lukas, T.J.; Van Eldik Mol. Cell. Biol. 12, 2359-2371, 1992

A:Title: Structure and expression of a calcium-binding protein gene contained within

A:Reference number: A44389; MUID:92236611

A:Accession: B44389

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1695-1906 <COL>

A:Cross-references: GB:M8284; NID:g212237; PIDN:AAB53767.1; PID:g212238

A:Accession: A44389

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1750-1906 <COL>

A:Cross-references: GB:M8283; NID:g211371; PIDN:AAA48647.1; PID:g211372

R:Yoshikai, S.I.; Ikebe, M. Arch. Biochem. Biophys. 29, 242-247, 1992

A:Title: Molecular cloning of the chicken gizzard telokin gene and cDNA.

A:Reference number: S28227; MUID:93073972

A:Accession: S28227

A:Molecule type: mRNA

A:Residues: 1750-1906 <YOS>

A:Cross-references: EMBL:M96655; NID:g212744; PIDN:AAA49083.1; PID:g212745

A:Accession: S78216

A:Molecule type: DNA

A:Residues: 1750-1906 <YOM>

A:Cross-references: EMBL:M96987

R:Olson, N.J.; Pearson, R.B.; Needleman, D.S.; Hurwitz, M.Y.; Kemp, B.E.; Means, A.R. Proc. Natl. Acad. Sci. U.S.A. 87, 2284-2288, 1990

A:Title: Regulatory and structural motifs of chicken gizzard myosin light chain kinas

A:Reference number: A35093; MUID:90192792

A:Accession: A35093

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 935-1438, 'O', 1440-1906 <OLS>

A:Cross-references: GB:M31048; NID:g212660; PIDN:AAA49069.1; PID:g212661

R;Guerriero Jr., V.; Russo, M.A.; Olson, N.J.; Putkey, J.A.; Means, A.R.  
 Biochemistry 25, 8372-8381, 1986  
 A;Title: Domain organization of chicken gizzard myosin light chain kinase deduced from a  
 A;Reference number: A25810; MUID:87157587  
 A;Accession: A25810  
 A;Molecule type: mRNA  
 A;Residues: 1258-1438, 'O', 1440-1906 <GUE>  
 C;Genetics:  
 A;Introns: 1735/3; 1779/1; 1819/1  
 C;Superfamily: myosin-light-chain kinase, nonmuscle; fibronectin type III repeat homolog  
 C;Keywords: alternative initiators; ATP; calmodulin binding; phosphoprotein; phosphotran  
 F;542-599/Domain: immunoglobulin homology <IMM1>  
 F;935-1906/Product: immunoglobulin homology <IMM2>  
 F;1098-1158/Domain: myosin-light-chain kinase, 108k, smooth muscle (from 5.5kb transcrip  
 F;1451-1708/Domain: immunoglobulin homology <IMM2>  
 F;1459-1467/Region: protein kinase ATP-binding motif  
 F;1750-1906/Product: telokin (kinase-related protein KRP) (from 2.7 kb transcript) #stat  
 F;1808-1869/Domain: immunoglobulin homology <IMM3>

## Query Match

2.3%; Score 145; DB 1; Length 1906;  
 Best Local Similarity 19.3%; Pred. No. 0.26;

Matches 179; Conservative 109; Mismatches 305; Indels 334; Gaps 45;

QY 140 PPSLRDTERIMSPSPKSCASACQOTGYOTRTRICLAEMVSLGSEASE-----GQ 193  
 DB 1111 PPAVY-----SWTIDSK-----AIKSSKSIIVISQEGILCSLTIEKVPEDGGE 1153  
 QY 194 H-CMGQDCTACDLTCPMGOVNAADCDACMCDPMLHGAVALPGAPASGAAIYLLTKTPKL 252  
 DB 1154 YKCAENNAAG-----KAEC-AC---KVLVEDTSSIRAKAPAKKTKKPPVTTLPV 1199  
 QY 253 LQTQDSGRFRIPGLCPDGKSLIKTKYKPAIVLTMPKTSLKATIKAEFVAETPYMV 312  
 DB 1200 LSTESSE-----ATVKKKP-----APKTPPKKAT-----PQGIT 1228  
 QY 313 MNPETKARRAGOSYSLCCKATGKRPDKYFWYHNDTLDDSLY-----KHESLVLRKL 366  
 DB 1229 QPFDKRVKRGESVELPAKAVG--TAPITCTWMKFRKQIQEVEYKIKENAESKLTITSST 1287  
 QY 367 QOHQAGEYFCKAOSDAGAVSKVAQLIVTASDE--TPC-NPVESYLIRLPHDFONAT 422  
 DB 1288 KOEHCGCYTLIVENKKGSRQAQVNLTVVDKPPDPAITPCASDIRSSL-----T 1336  
 QY 423 NSFY--YDVGR-----PVKTCAG-----QDNGICRDVAQNC 454  
 DB 1337 LSWGSSYDGSNAVQSYTVEIWNVDNKTDLTTCRSTSFNVOLDADREKFEVRAANV 1396  
 QY 455 CGIS-----TTEEREIOCGSYTLPTKVAKCSQORCTETSIYRGRVSADNG 502  
 DB 1397 YGISEPQSEEVYKVGKEQEBELAEELSDDEGKETEVNRTVTIN--TEQKVSVDYNI 1455  
 QY 503 EPM---RFQHVYVGNRSVMTGKGT-----TLHVPODTERLV 538  
 DB 1456 EHRIGSKFQGVPRVLVETKTKGKVAAGKFEKAYSAKKEKNIRDETSIMCLHHR-----L 1510  
 QY 539 LTFVDRLQKQFVNTKYLPRNKKGSAAVFEIKMLRKKEPITLEAMETNIIPLGVEVGEDPM 598  
 DB 1511 VQCDAAEEKANIYVLEM--VSGGELFERI-----IDED-- 1543  
 QY 599 AELEIPR---SFYRQ--NRPYIGKAKASYTLDPNINISTATAQDNLFINDEGDTFP 653  
 DB 1544 -FELTERECIKYMRQISEVEYIHKGIVHLDPKINIMCVNKTGTSIKLID----- 1594  
 QY 654 LRTVGMSEVDFRDEVTSEPLNAGKVYVHLDTQYKKEPHS-----TVKLMS----- 700  
 DB 1595 -----FGLARRLE-----SAGSLKVLFGTPEFVADVINYEPIGETDMKMSGVICYI 1642  
 QY 701 -----LNDPTGLWEE-----GDFKFNQRRNK--REDRTFLVGNLEIRRRLLFNLDVPE 748  
 DB 1643 LVSGLSLSPFMDNDNETLANVTSATWDFDEAFDEISDADDFISN-----LTKKMKMS 1695  
 QY 749 SRRC-----FAKVRAYRSEERLPSQIQGVVIVINLEPRGTGLSNPR 791

DB 1696 RLNTQCLQHPWLQDKTRNMEAKKLSKDMKKYVARRKWO-----KFGHA--VR 1742  
 QY 792 AMGFEDS--YITSPNGACVAFACDDQSPDAYSAVYLAISLAGEIQAVNESSPKFNPAINIGV 849  
 DB 1743 AIGRLSSMAITSGMSGR-----KAGSSP----- 1766  
 QY 850 POPYLNKLYNRDDEHEDPRVAKTAFOJISMAKPRENSAESNGPIYAFENLRACEAPPSA 909  
 DB 1767 -----TSPINDKVENDEAFLEEVAEKPRHKP-----YFKTITLDEMEVVGSA 1810  
 QY 910 AHREFYQIEGDY--DYNTVPPNEDDPM 935  
 DB 1811 ARFD-CRIEG--YPDPEVMYKDDQPV 1834

## RESULT 12

JN0583  
 myosin-light-chain kinase (EC 2.7.1.117) 155k protein - bovine

C;Species: Bos primigenius taurus (cattle)

C;Date: 03-Feb-1994 #sequence\_revision 03-Feb-1994 #text\_change 20-Apr-2000

C;Accession: JN0583; P0490

R;Kobayashi, H.; Inoue, A.; Mikawa, T.; Kuwayama, H.; Hotta, Y.; Masaki, T.; Ebashi,

J. Biochem. 112, 786-791, 1992

A;Title: Isolation of cDNA for bovine stomach 155kDa protein exhibiting myosin light

A;Reference number: JN0583; MUID:93203148

A;Accession: JN0583

A;Molecule type: mRNA

A;Residues: 1-1176 <KOB>

A;Cross-references: GB:557131; NID:g2296838; PIDN:AAB25794.1; PID:g2296839

A;Experimental source: Stomach

A;Accession: P0490

A;Residues: 44-55;721-728;828-851;1002-1019 <KO2>

C;Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; fibronectin type I

C;Keywords: ATP; calmodulin binding; phosphotransferase

F;370-430/Domain: immunoglobulin homology <IMM1>

F;723-980/Domain: protein kinase homology <KIN>

F;731-739/Region: protein kinase ATP-binding motif

F;1083-1144/Domain: immunoglobulin homology <IMM2>

## Query Match

2.3%; Score 144.5; DB 2; Length 1176;  
 Best Local Similarity 20.3%; Pred. No. 0.14;

Matches 179; Conservative 115; Mismatches 306; Indels 281; Gaps 47;

QY 166 GQGVQVOTRTRICLAEMVSLGSEASEGQHCMDG-----CTACDLTCPMGOVNAADCDACM 220  
 DB 393 GKT-LKTRKTVLSQESLSYSTIEK---ALPEDRGILYKVAKN---SAGQAESSCVY-- 443  
 QY 221 QDFMLHGAVALPGAPASGAATYLLTKTPKLLTQTDSDGRFRIPGLCPDGKSLIKITKV 280  
 DB 444 -----TVDP--DAPTSNA-----KADEMKAARRKSSLPVLTGESP-----ATV 482  
 QY 281 KPAIVLTMRKTSLSKAATTAEFVRAETPYVMANPETKARRAGOSYSLCCATGKPRPK 340  
 DB 483 KKKP---AKTTPPKAM-----PQIIQFPEDQKVRAGESVELFQKAVAG--TQPIR 528  
 QY 341 YFWV--HNDTLDDSLYKHE-----SKLVLRKLOHQAGEFCKAQSAGAVKVAQNLIV 394  
 DB 529 CTWKKFRKQIQDSBHIKVENSEQSKLTITRAAOEHCGCTYLLVENKLSGRQAQVNLTV 588  
 QY 395 TASDE--TPC-NPVESYLIRLPHDFONATNSFY--YDVGRCPVKTCAGQDNGIRC 447  
 DB 589 DKDPDPACTGCASDIRSSL-----TISWISSIDG-----SAVQISVET 630  
 QY 448 RDAVQ-----NCCGIS-----KTEEREIOCGSYTLPTKVAKCSQORCTET 488  
 DB 631 WDSVDKTKWELACRSTSFNVODLLPDREYKFRVRAINVCSTSEPSQESLTLALGEKPEE 690  
 QY 489 RSLYRGRVSAADNGEPARFPHYVYMGNSRVSMTGKGTFTLHVQDTERLVLTVD--RL 545  
 DB 691 EPDDEVESDDDERP-----EVDYR--TVTV-----NTEQKVSDEYDIEERL 731



F:151,316,330,457,584,710,1069/Binding site: carbohydrate (Asn) (covalent) #status predi  
 F:167-226/Disulfide bonds: #status predicted  
 F:266,270/Disulfide bonds: interchain #status predicted  
 F:612/Modified site: erythro-beta-hydroxyasparagine (Asn) #status predicted

Query Match 2.3%; Score 143; DB 1; Length 1172;  
 Best Local Similarity 20.8%; Pred. No. 0.18;  
 Matches 105; Conservative 56; Mismatches 189; Indels 154; Gaps 29;

QY 86 CAPRLRLRLA-----TMDTASGTVGVHSPRGFMCLNRQRPQ 128  
 DB 360 CASPFVEGSCCSCHSVDEBWSWPAEMTQCSVT-----CGSGTQGR 406  
 QY 129 NCSNYTRFLCP-----PGSLRDLTERI-----WSPWSPWKSACAGQTVQTRICL 178  
 DB 407 SCVTSNTCLGPIQLTRACLSKODFRIQDDGSHWSPWSSCVTCG-VGNITRILCN 465  
 QY 179 AEWVSL---CSEASEGQHMGQDC-----TACDLTCPMG--QVNAACD--- 217  
 DB 466 SPVPOMGKMKSGRETAKOGAPCPIDGRWSPWSPWSPWACTVTCAGIRERTRVCSPE 525  
 QY 218 -----AC-----MC--ODFMLHGAVSLP--GAPASGAIVLTKTKPKLLTQDSD 259  
 DB 526 PQYGRACVGDVDERQMKRSCPVDCLSNCPGPGACS-----SFPD 569  
 QY 260 GRFRIPGLCP-----DGKSLIKITKVFAPIVLTPMTSLKAATIKAEFVRAETP--YVW 312  
 DB 570 GSWSC-GFCVVGFLNGTHCEDDECALVPDICEFSTKVPKVCNTQPGFHLPCPPRYRG 628  
 QY 313 MNPETARRRAGQSVSLCCKATGKPRDKYFYVINDTLDPSTLYKHESKLVLRKLQHQHAG 372  
 DB 629 NOPVGVGLEAKTEKQVCEPE-NPCKDK---TNN-----CHKHACIYLGHFSDPM-- 675  
 QY 373 EYFCKAQSADGAVKSKAQLIVTASDETPCNVPESYLIRLPHDCFQNAFNSFYVDY-GR 431  
 DB 676 -YKCECQTG-----YAGDELICG--EDSDLDGPMNLNLYCATNATYHCTIKDN 719  
 QY 432 CPVKTCAGQDNGIRCDAYVONCC-----GISKTEEREIQCSTGYLLPTVAKESG--- 481  
 DB 720 CPHLPNNGGDEFD--KDGIGDACDDDDDDNDGY--TDEKD-NCQLLFNPRQADYDXDEVG 773  
 QY 482 --CORCTETRSYRGRVSAADNGE 503  
 DB 774.DRCDNCPYVHN--PAQIDTDNNGE 795

## RESULT 15

T33922  
 Hypothetical protein Y8A9A.2 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
 C:Accession: T33922  
 R:Courtney, L.; Langston, Y.; Drone, K.; Mead, K.  
 submitted to the EMBL Data Library, February 1999  
 A:Description: The sequence of C. elegans cosmid Y8A9A.  
 A:Reference number: Z21439  
 A:Accession: T33922  
 A:Status: preliminary;  
 A:Molecule type: DNA  
 A:Residues: 1-1360 <CON>  
 A:Cross-references: EMBL:AF125461; PIDN:AAD12852.1; GSPDB:GN00020; CESP:Y8A9A.2  
 A:Experimental source: strain Bristol N2; clone Y8A9A  
 C:Genetics:  
 A:Gene: CESP:Y8A9A.2  
 A:Map position: 2  
 A:introns: 43/3; 69/3; 134/3; 174/1; 604/2; 643/3; 718/3; 787/3; 848/2; 944/2; 1044/2; 1

Query Match 2.3%; Score 143; DB 2; Length 1360;  
 Best Local Similarity 26.1%; Pred. No. 0.22;  
 Matches 61; Conservative 16; Mismatches 95; Indels 62; Gaps 11;

QY 54 GEMTTLFNT-DYDGKG-DYERLDAIRFYVG-----DRVCARPLRLRLAETDWT-- 100  
 DB 1108 GEMVSVSGNDTCGSCGQETRRKRKCLSLQYGCACGTGNATDTVCASSVCLFPRTSCTGF 1167  
 QY 101 --PAGSTGVVHSPRGFMCLNRQRPQNGCSNYTRFLCPPGSLRDLTERIWSWSPW 158  
 DB 1168 KKNVNTIGRTFYCGPLFVPAFNPQ---TTC-----CDP-----EXTGLMDMGAW 1211  
 QY 159 SKCSAAGQTVQTRFTRICLAEWVSL-CSEASEGQHMGQDCCTACDLTCPMGVNACD 217  
 DB 1212 TTCSATCGGCGTQTRSRKTCASAPYGCCTGDDLTETQSCAKQVCTTGACCCAGKFAVATGYD 1271  
 QY 218 -ACMCD-----FMLHGAIV-----SLPGAPASGA 241  
 DB 1272 GAQYCDNTPREVCTGTWTEWATLEGAVCNDTCGNCGLIPTSRYCFPSGCGCSGA 1325

Search completed: October 9, 2001, 11:50:42  
 Job time: 227 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 9, 2001, 11:50:15 ; Search time 14.83 Seconds

(without alignments)  
2734.895 Million cell updates/sec

Title: US-09-609-383-2

Perfect score: 6317  
Sequence: 1 MVTGKAWFSEFLVEVTSVL.....QSGVVASLRFPVRAQOPLIN 1184

Scoring table: BLOSOM62  
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	170	2.7	890	1	AT58_HUMAN
2	169	2.7	1170	1	P35441 mus musculu
3	166	2.6	1170	1	TSPL_BOVIN
4	162	2.6	1170	1	TSPL_BOVIN
5	161	2.5	1694	1	SN_MOUSE
6	160	2.5	905	1	AT58_MOUSE
7	159.5	2.5	2569	1	LMA3_MOUSE
8	157	2.5	1170	1	TSPL_HUMAN
9	151.5	2.4	4393	1	PGBM_HUMAN
10	146	2.3	1239	1	NRG_DROME
11	145	2.3	1173	1	TSPL_XENLA
12	145	2.3	1906	1	KMLS_CHICK
13	144.5	2.3	1176	1	KMLS_BOVIN
14	144.5	2.3	1914	1	KMLS_HUMAN
15	143	2.3	1172	1	TSPL_HUMAN
16	142.5	2.3	4544	1	LRPI_HUMAN
17	142	2.2	1074	1	SM5A_HUMAN
18	140.5	2.2	2481	1	UN52_CAEEL
19	139.5	2.2	1056	1	MUC5_HUMAN
20	139	2.2	1122	1	TIE2_MOUSE
21	139	2.2	1522	1	BA13_HUMAN
22	137.5	2.2	1493	1	NR01_MOUSE
23	136	2.2	2907	1	FBN2_MOUSE
24	135	2.1	1147	1	KMLS_RABIT
25	135	2.1	2911	1	FBN2_HUMAN
26	134.5	2.1	1450	1	MPSE_CHICK
27	133.5	2.1	610	1	TOH2_CAEEL
28	132	2.1	1178	1	TSPL_CHICK
29	132	2.1	1211	1	AT52_HUMAN
30	131.5	2.1	3707	1	PGBM_MOUSE
31	130.5	2.1	1125	1	TIE2_BOVIN
32	130	2.1	470	1	PROP_CAVPO
33	130	2.1	867	1	SSPO_BOVIN

34	130	2.1	1572	1	BA12_HUMAN	060241 homo sapien
35	130	2.1	2437	1	NOTC_BRARE	P46530 brachydanio
36	129.5	2.1	967	1	AT51_HUMAN	Q9uh18 homo sapien
37	129.5	2.1	1077	1	SM5A_MOUSE	Q62217 mus musculu
38	129.5	2.1	1093	1	SM5B_MOUSE	Q60519 mus musculu
39	129	2.0	2029	1	LAR_DROME	P16621 drosophila
40	128.5	2.0	807	1	FSP0_RAT	P35446 rattus norv
41	128.5	2.0	1172	1	TSPL_MOUSE	Q03350 mus musculu
42	128.5	2.0	1338	1	VGR1_HUMAN	P17948 homo sapien
43	128.5	2.0	1584	1	BA11_HUMAN	Q14514 homo sapien
44	128	2.0	997	1	AT57_HUMAN	Q9ukp4 homo sapien
45	127.5	2.0	1205	1	AT52_BOVIN	P79331 b adam-ts 2

## ALIGNMENTS

```
RESULT 1
ID      AT58_HUMAN      STANDARD:      PRT:      890 AA.
AC      Q9UP79; Q9NZS0;
DT      01-OCT-2000 (Rel. 40, Created)
DT      01-OCT-2000 (Rel. 40, Last sequence update)
DT      01-OCT-2000 (Rel. 40, Last annotation update)
DE      ADAM-TS 8 PRECURSOR (EC 3.4.24.-) (A DISINTEGRIN AND METALLOPROTEINASE
DE      WITH THROMBOSPONDIN MOTIFS 8) (ADAMTS-8) (ADAM-TS8) (METH-2) (METH-8).
GN      ADAMTS8 OR METH2.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Lung;
RX      MEDLINE=99367466; PubMed=10438512;
RA      Vazquez F., Hastings G., Ortega M.-A., Lane T.F., Olkemus S.,
RA      Lombardo M., Iruela-Arispe M.L.;
RT      "METH-1, a human ortholog of ADAMTS-1, and METH-2 are members of a new
RT      family of proteins with angio-inhibitory activity.";
RL      J. Biol. Chem. 274:23349-23357(1999).
RN      [2]
RP      SEQUENCE OF 195-440 FROM N.A.
RC      MEDLINE=20079168; PubMed=10610729;
RX      Georgiadis K.E., Hirohata S., Seidlin M.F., Apre S.S.;
RA      "ADAM-TS8, a novel metalloprotease of the ADAM-TS family located on
RA      mouse chromosome 9 and human chromosome 11.";
RT      Genomics 62:312-315(1999).
RL      -1- FUNCTION: HAS ANTI-ANGIOGENIC PROPERTIES.
CC      -1- COFACTOR: BINDS ONE ZINC ION (BY SIMILARITY).
CC      -1- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR
CC      MATRIX (BY SIMILARITY).
CC      -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN ADULT AND FETAL LUNG, LOWER
CC      EXPRESSION IN BRAIN, PLACENTA, HEART, STOMACH AND FETAL BRAIN AND
CC      KIDNEY.
CC      -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE 1 DOMAINS ARE IMPORTANT
CC      FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC      -1- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY
CC      SIMILARITY).
CC      -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B (ZINC
CC      METALLOPROTEASE); ALSO KNOWN AS THE REPROLYSIN SUBFAMILY.
CC      -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC      -1- SIMILARITY: CONTAINS 2 TSP TYPE-1 DOMAINS.
-----
CC      THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      EMBL; AF060153; AAD48081.1; -
CC      EMBL; AF175283; AAF25806.1; -
DR
```

DR HSSP: P34179; 11AG.  
 DR MIM: 605175; -.  
 DR InterPro: IPR000130; -.  
 DR InterPro: IPR000884; -.  
 DR InterPro: IPR001590; -.  
 DR InterPro: IPR002870; -.  
 DR Pfam: PF000090; tsp.1; 2.  
 DR Pfam: PF01421; Repolysin; 1.  
 DR Pfam: PF01562; Pep\_M12B\_propep; 1.  
 DR PROSITE: PS50215; ADAM\_MEPRO; 1.  
 DR PROSITE: PS00142; ZINC\_PROTEASE; 1.  
 DR PROSITE: PS50092; TSP1; 2.  
 DR PROSITE: PS00427; DISINTEGRINS; FALSE\_NEG.  
 DR Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;  
 DR Repeat; Extracellular matrix; Heparin-binding.  
 FT SIGNAL 1 27  
 FT PROPEP 28 214  
 FT CHAIN 215 890  
 FT METAL 364 364  
 FT ACT\_SITE 365 365  
 FT METAL 368 368  
 FT METAL 374 374  
 FT DOMAIN 439 526  
 FT DOMAIN 527 583  
 FT DOMAIN 584 690  
 FT DOMAIN 691 832  
 FT DOMAIN 833 890  
 FT DOMAIN 202 205  
 FT CARBOHYD 345 345  
 FT CARBOHYD 401 401  
 FT CARBOHYD 466 466  
 FT CARBOHYD 491 491  
 FT CARBOHYD 600 600  
 FT CONFLICT 195 195  
 FT CONFLICT 413 440  
 SEQUENCE 890 AA: 96671 MW: 57070 EED03D5739D3 CRC64;  
 IHEKLYCKGSGDCLDAPALPLPTGL -> FSGCHLQGM

Query Match 2.7%; Score 170; DB 1; Length 890;  
 Best Local Similarity 21.4%; Pred. No. 0.00044;  
 Matches 110; Conservative 49; Mismatches 224; Indels 132; Gaps 25;

QY 50 LESBGEWTLFNDYDGGKDYERLDAIRYGD--RVCAKRLRLKRTDWTDPAGSTGQ 107  
 DB 427 LDAPGALPL-PTGLPGRMALYOLDQOCROIFPDRFHCHNTSAODVCAOLMCHTGAEP 485  
 QY 108 VHGSPREGFWCLNREORPGONCSNTYVRLCPGSLRDRTERI-----NSPWSPMK 160  
 DB 486 LGHTKNGSLPMAGCTPGCGHLCSEGS---CLP---EEVEVRKPRVVDGGMARPMGPWGE 538  
 QY 161 CSAACGQTVQVTRTRICLAEMVSLCSEASEGQCMG---ODCTACDLTCEPMGOVNAD 215  
 DB 539 CSRTCG-GGVQFSHRECKD-----PEPONGRHVCLGRRAKYOSCHTEE--CPPD--GKS 587  
 QY 216 CDACMCDQDEFLHVAVSLPGAPASGAAYILITPKILITOTDSDGRIRTPGLC-PDGKSI 274  
 DB 588 FREQQCEKNAYVYTDMDG-----LLQVPRYAGVSPRD---RCKLFRANRGRSE 635  
 QY 275 LKTKKFAPIVLTMRKTSIKATIRAEFRATPYVMNPE--TKARRAGQSVSLCKCA 332  
 DB 636 FVFEAKVIDGLCPET--LAICVAGQCVKACGDHVDSPRLDKGCGVGGGNGSCRKY 693  
 QY 333 TGNPRDKFWIYND-----TLID-----PSLYKHSKVLKRLQHQAGEYFCKA 378  
 DB 694 SGLTFTNY--GYNDIVTTPAGATNIDVKORSHPGVQNDGNALYATAD---GQYLING 747  
 QY 379 QSDAGAVKS-----KYAQLIVASDETPCNPPESTLIL---PHCF-----418  
 DB 748 NLAISAEODILVKGITLTKSGSIALERLQSRPLPEPLTVGLTVPGGVPPKVKYTF 807  
 QY 419 -----ONATNSFY-----DVGRCFVKTAGOODNIRPRD---A 450

DB 808 FVNDVDFSMOSSKERATTNIIOPLLHAQWVLDWSECSSTCGAGMORTVECRDPSGQA 867  
 QY 451 VONCCGISKTEEREIQCSTGLTPTKVAKECCORC 485  
 DB 868 SATCNKALKPED-----AKPESQLC 888  
 RESULT 2  
 TSP1\_MOUSE STANDARD; PRT: 1170 AA.  
 ID TSP1\_MOUSE  
 AC P35441;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE THROMBOSPONDIN 1 PRECURSOR.  
 GN THBS1 OR TSP1.  
 OS Mus musculus (mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_Taxid=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92128941; PubMed=1774063;  
 RA Lawler J., Duquette M., Ferro P., Copeland N.G., Gilbert D.J.,  
 RA Jenkins N.A.;  
 RL "Characterization of the murine thrombospondin gene.";  
 RL Genomics 11:587-600(1991).  
 [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92147683; PubMed=1371115;  
 RA Laherty C.D., O'Rourke K., Wolf F.W., Katz R., Seldin M.F.,  
 RA Dixit V.M.;  
 RL "Characterization of mouse thrombospondin 2 sequence and expression  
 RL during cell growth and development.";  
 RL J. Biol. Chem. 267:3274-3281(1992).  
 [3]  
 RP SEQUENCE OF 1-490 FROM N.A.  
 RX MEDLINE=9037546; PubMed=2398070;  
 RA Bornstein P., Alf D., Devarayalu S., Framson P., Li P.;  
 RL "Characterization of the mouse thrombospondin gene and evaluation of  
 RL the role of the first intron in human gene expression.";  
 RL J. Biol. Chem. 265:16691-16698(1990).  
 CC -1- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND  
 CC CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN,  
 CC LAMININ AND TYPE V COLLAGEN.  
 CC -1- SUBUNIT: HOMOTRIMER, CROSS-LINKED BY DISULFIDE BONDS.  
 CC -1- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.  
 CC -1- SIMILARITY: CONTAINS 1 VWFC DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 7 TSP TYPE-1 DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 7 TSP TYPE-3 DOMAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions, as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M62470; AAA50611.1;  
 DR EMBL: M62450; AAA50611.1; JOINED.  
 DR EMBL: M62451; AAA50611.1; JOINED.  
 DR EMBL: M62452; AAA50611.1; JOINED.  
 DR EMBL: M62453; AAA50611.1; JOINED.  
 DR EMBL: M62454; AAA50611.1; JOINED.  
 DR EMBL: M62455; AAA50611.1; JOINED.  
 DR EMBL: M62456; AAA50611.1; JOINED.  
 DR EMBL: M62457; AAA50611.1; JOINED.  
 DR EMBL: M62458; AAA50611.1; JOINED.  
 DR EMBL: M62459; AAA50611.1; JOINED.  
 DR EMBL: M62460; AAA50611.1; JOINED.  
 DR EMBL: M62461; AAA50611.1; JOINED.



OY	59		-----LFTNDVGGGKGDVERLDIAIRFYGGRVCAARPLREARTTD	98
Db	338	ECHCONSNTATICKKYSCPTMPCSNATVPDGE	-----CCPKCMPSDSADOG	381
OY	99	WTPAGSTGYVVHGSBREGFW	-----CLANRBPQGNCSNVTYRLCPGSLRRDTERI	151
Db	382	WSPWSB	-----WTSGCATCGNGIDQRGRSDLSLNRR--CEGSSVQTRTCHIQ	426
OY	152	-----WSPWSRSKSCSAACGTGYQTRTICLAEMVSL--CSEASEGOHCM	196	
Db	427	ECDKRERKODGGSHSWPSSCSYVCBD-GVITRIIRLCNSPDPOMMGKPCDEEAPRTRAK	485	
OY	197	GDDC-----TACDLTC-----PMGVNA-DC-----PACMKOD	223	
Db	486	KDACPRLNGMRPWMDICSVTGGGQVRSRSLCNPNPTQPGFKCKCVQDYDENOVCKKD	545	
OY	224	FMLHGAVSLPGGAPASGAATLYLLTKPRKLLTDTSDGRFRIPGLCPDGSKSLIKTKYKFA	283	
Db	546	CPIDGCLSNPCFAGAK-----CTSYD-----DGSWK-C-GACPGRYS	580	
OY	284	PVLVIMPTKTSLKAAITKAEPBATETPYMANNPETKARRRGOSVSLCKATG--KRPDKY	341	
Db	581	-----GNGIQ--CKDVCEKEVPDAC	595	
OY	342	FWYHNDTLDPISLYKHESKLVRKLDQHAGEFYFCRAQSADG-----	383	
Db	600	F-NHN-----GEHRCK-RNTPRYNCILPCRPRRTGSOPGR	633	
OY	384	AVKSRYAQLVTASDETPCNPVPESYLIRPHDFQGNATNSF---YYD-VGCRPVKTCAG	435	
Db	633	GVEHAMAKQVC---KPRNCTDG-----THDCKNKKANCNYLGHSDBMYRWCECK--PG	681	
OY	440	QQDNIGRC	447	
Db	682	YAGNGIIC	689	
RESULT	3			
TSP1_BOVIN		STANDARD;	PRT; 1170 AA.	
ID	Q28178; Q28179;			
DT	01-NOV-1997 (Rel. 35, Created)			
DT	01-OCT-2000 (Rel. 40, Last sequence update)			
DT	01-OCT-2000 (Rel. 40, Last annotation update)			
DE	THROMBOSPONDIN 1 PRECURSOR.			
GN	THBS1 OR TSP1 OR TSP-1.			
OS	Bos taurus (bovine).			
OS	Eukaryota; Metazoa; Chordata; Cranialta; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;			
OC	Bovidae; Bovinae; Bos.			
OX	NCBI_TaxId=9913;			
RN	(1)			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-HOLSTEIN; PUBSUB-Tooth;			
RC	MEDLINE=9813773; PubMed=9507054;			
RC	Ueno A., Yamashita K., Nagata T., Tsurumi C., Miwa Y., Kitamura S.,			
RA	Inoue H.,			
RA	"cDNA cloning of bovine thrombospondin 1 and its expression in			
RT	odontoblasts and predermal."			
RT	Odontoblasts.			
RL	Biochim. Biophys. Acta 1382:17-22(1998).			
RN	[2]			
RP	SOURCE OF 1-18 AND 710-1170 FROM N.A.			
RC	TISSUE-Aortic endothelium.			
RA	Zafar R.S., Moll Y.D., Womack J.F., Walz D.A.;			
RL	Submitted (JUN-1995) to the EMBL/GenBank/DBD databases.			
-1-	FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND			
CC	CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN			
CC	LAMININ AND TYPE V COLLAGEN. MAY PLAY A ROLE IN DENTINOGENESIS			
CC	AND/OR MAINTENANCE OF DENTIN AND DENTAL PULP.			
-1-	SUBUNIT: HOMOTRIMER, CROSS-LINKED BY DISULFIDE BONDS.			
-1-	TISSUE SPECIFICITY: ODONTOBLASTS.			
-1-	SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.			



RL Submitted (May-1995) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND  
 CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN,  
 CC LAMININ AND TYPE V COLLAGEN.  
 CC -1- SUBUNIT: HOMOTRIMER, CROSS-LINKED BY DISULFIDE BONDS.  
 CC -1- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.  
 CC -1- SIMILARITY: CONTAINS 3 WFEC DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 7 TSP TYPE-3 DOMAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: X96540; CAA65385.1; -  
 DR EMBL: X87620; CAA60952.1; -  
 DR InterPro: IPR000561; -  
 DR InterPro: IPR000884; -  
 DR InterPro: IPR001007; -  
 DR InterPro: IPR001881; -  
 DR Pfam: PF00008; EGF; 2.  
 DR Pfam: PF00090; TSP-1; 3.  
 DR Pfam: PF00093; WFC; 1.  
 DR PROSITE: PS00022; EGF\_1; FALSE\_NEG.  
 DR PROSITE: PS01186; EGF\_2; 1.  
 DR PROSITE: PS50092; TSP1; 3.  
 DR PROSITE: PS01208; WFC; 1.  
 KW Glycoprotein; Cell adhesion; Calcium-binding; Heparin-binding; Repeat;  
 KM EGF-like domain; Signal.  
 FT SIGNAL 1 18  
 FT CHAIN 19 1170  
 FT DOMAIN 19 232  
 FT DOMAIN 318 375  
 FT DOMAIN 379 430  
 FT DOMAIN 435 491  
 FT DOMAIN 492 546  
 FT DOMAIN 547 587  
 FT DOMAIN 588 645  
 FT DOMAIN 646 690  
 FT DOMAIN 723 758  
 FT DOMAIN 759 781  
 FT DOMAIN 782 817  
 FT DOMAIN 818 840  
 FT DOMAIN 841 878  
 FT DOMAIN 879 914  
 FT DOMAIN 915 950  
 FT DOMAIN 951 1170  
 FT SITE 926 928  
 FT SITE 926 928  
 FT DISULFID 270 270  
 FT DISULFID 551 562  
 FT DISULFID 556 572  
 FT DISULFID 575 586  
 FT DISULFID 592 608  
 FT DISULFID 617 644  
 FT DISULFID 620 663  
 FT DISULFID 650 676  
 FT DISULFID 657 676  
 FT DISULFID 676 689  
 FT CARBOHYD 151 151  
 FT CARBOHYD 316 316  
 FT CARBOHYD 330 330  
 FT CARBOHYD 455 455  
 FT CARBOHYD 582 582  
 FT CARBOHYD 708 708  
 FT CARBOHYD 936 936  
 FT CARBOHYD 1067 1067  
 FT CONFLICT 535 535  
 A -> V (IN REF. 3).

FT CONFLICT 748 748 S -> T (IN REF. 3).  
 SQ SEQUENCE 1170 AA; 129862 MW; 9CF1F5F589A051 CRC64;  
 Query Match 2.68; Score 162; DB 1; Length 1170;  
 Best Local Similarity 21.6%; Pred. No. 0.0024;  
 Matches 102; Conservative 43; Mismatches 153; Indels 174; Gaps 25;  
 QY 86 CARLRLEART-----TDWPPAGSTGVGVHSPREGFWCINLRBQPGNC 130  
 DB 360 CADPWEVEGCCPSCVHDEGEGMSPMAEWTECSAT-----CGSGTQQRHSC 406  
 QY 131 SNVYVRFCLP-----PSLRDTERI-----WSPWSPKSCAACGOTVOTRRICIAE 180  
 DB 407 DVTNTGCGPDIQIRACSLGKCDHRIRQDGGWSHMSPPSSCYCG-VGNTIRLRICNSP 465  
 QY 181 MVSL-----CSEASEGGHCHMGQDC-----TACDITCPMGQVYNADCCACMCODF 224  
 DB 466 VPOMGRSCSKSGRETRACQGPDPVDGRWSPWSPMSACTVTC-AGGIRERTVQNSPSE 523  
 QY 225 MLHGAVSLPGGA-----PAGCAIYLILKT--PKLLTQTDSDGRFRIRIGLCPDG 271  
 DB 524 P0HGGKDCVGGAKDQMCNRRKCPIDG---LSNPCEPAGECSSFPDGSWSC-GSCPQG 578  
 QY 272 -----KSILKITKVFAP-----IYLTMP-----KTSLKAT 298  
 DB 579 FLNGCTHCEIDDECAVYTVDFATSKAHRCVNTNPGHICLPFRYKGNDFGYLEAR 638  
 QY 299 IKAEFVRAEPFYVMWNPETKARRAGOSVSLCKATGKRPDPKYWYHNDLLDPSLYKHE 358  
 DB 639 TEKQVCEPENE-----CK-----DK-----TSHCRHA 661  
 QY 359 SKVLRLKIQHQAEYFCKKASDAGAVKSKVAQLIVASDETPCNPPESTYLIRLPDCE 418  
 DB 662 ECITLGHFSDPM---YKCECQTG-----YAGDLIGG-EDSLDGMPPKNL 703  
 QY 419 QNATNSFYDYD-GRCPVKTCAGQODNGIRCDAYONCC-----GISKTEE 462  
 DB 704 VCAINATYHCYKDKNCPLPNSGQEDFD---KDIGDACDDDDNDGVSDEKD 752  
 RESULT 5  
 ID SN\_MOUSE STANDARD: PRT; 1694 AA.  
 AC 062230; 055216; 062228; 062229;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DE 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE STALOADESIN PRECOURSOR (SER).  
 GN SN OR SA.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_Taxid=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC STRAIN-CS7BL/6; TISSUE-Macrophage;  
 RX MEDLINE=9500950; PubMed=7925291;  
 RA Crocker P.R., Mucklow S., Boukson V., McWilliam A., Willis A.C.,  
 RA Gordon S., Milon G., Kelm S., Bradfield P.;  
 RT "Stalodhesin, a macrophage stialic acid binding receptor for  
 RT haemopoietic cells with 17 immunoglobulin-like domains.";  
 RL EMO J. 13:4490-4503(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98051930; PubMed=9383289;  
 RA Mucklow S., Gordon S., Crocker P.R.;  
 RT "Characterization of the mouse stalodhesin gene, Sn.";  
 RL Mamm. Genome 8:934-937(1997).  
 RN [3]  
 RP CHARACTERIZATION.  
 RC STRAIN-CS7BL/6; TISSUE-Spleen;  
 RX MEDLINE=91266893; PubMed=2050106;



OY	262	FRIG-----LCPOGSKILKTIKVRKAPILVMPK-----SIKATI	299
Db	252	NILGDPVTLTCRVNSYIPAVSAVQMRDGVNLGVTGHLRLFSAAAMDSCGYTQATND	311
OY	300	KAEPVRAETPYVMANPEPTKARRAG-----QSVSLCKAKGKPRD--KYFWYHNDTLDP	352
Db	312	MGLSVSSPLSLIHYPMAYKKNPPAGPVLNENYTLIC-STPKAPDOLKRWIKHILLD	370
OY	353	SLYHESKLVLRKLOOHAGEYFCCKAOSDAGAVKSKVAQLIVTASDETP-----	401
Db	371	A--HASTLHLPRATRADTGFYFCGVQNMGSRRSSPLSVVRYRPLRDLTTFLETQAG	427
OY	402	-----CNPPEST-LIRLPHDFONATNSFYIDVGRCPVTKCAGQODNGIRCDAYONC	454
Db	428	LVGILHCVSVSEPLATVVLSHGGLTLAUNS-----GENDFNPRFR--ISSA	471
OY	455	CGISIKTERERELQCSGYLLPFTKVAKECSQRCQCTFSTRIVGRVSAADNEPMRFGVIVYGN	514
Db	472	PNSLRLEIRDO-----PADSGEYTCCLAVNSLGN	500
OY	515	SRVSMGTGYKGFETHVPQDE	535
Db	501	STSSLDIFYANVARLLINPSAE	521

RESULT	6
AC	AT58_MOUSE
ID	AT58_MOUSE
STANDARD:	PRT: 905 AA.
AC	P57110.
DT	01-OCT-2000 (Rel. 40, Created)
DT	01-OCT-2000 (Rel. 40, Last sequence update)
DT	01-OCT-2000 (Rel. 40, Last annotation update)
DE	ADAM-TS 8 PRECURSOR (EC 3.4.24.-) (A DISINTEGRIN AND METALLOPROTEINASE WITH THROMBOSPONDIN MOTIFS 8) (ADAMTS-8) (ADAM-TS8) (METH-2).
GN	ADAMTS8.
OS	Mus musculus (Mouse).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX	NCBI_Taxid:10090;
RP	[1]
RF	SEQUENCE FROM N.A.
RX	MEDLINE-20079168; PubMed-10610729;
RA	Georgiadis K.E., Hirohata S., Seidlin M.F., Apte S.S.;
RT	"ADAM-TS8, a novel metalloprotease of the ADAM-TS family located on mouse chromosome 9 and human chromosome 11.";
RL	Genomics 62:312-315(1998)
CC	-1- FUNCTION: HAS ANTI-ANGIOGENIC PROPERTIES (BY SIMILARITY).
CC	-1- COPECTOR: BINDS ONE ZINC ION (BY SIMILARITY).
CC	-1- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR MATRIX (BY SIMILARITY).
CC	-1- TISSUE SPECIFICITY: EXPRESSED SPECIFICALLY IN ADULT LUNG AND HEART AND LOW EXPRESSION DURING MOUSE DEVELOPMENT.
CC	-1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE 1 DOMAINS ARE IMPORTANT FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC	-1- P1M: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY SIMILARITY).
CC	-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B (ZINC METALLOPROTEASE); ALSO KNOWN AS THE REPROLYSIN SUBFAMILY.
CC	-1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC	-1- SIMILARITY: CONTAINS 2 TSP TYPE 1 DOMAINS.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb.ch/announce/">http://www.isb.ch/announce/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).
CC	-----
DR	EMBL; AF175282; AAF25805.1; -
DR	MGD; MGI:1353468; Adamts8.
DR	InterPro; IPR000130; -
DR	InterPro; IPR000884; -

DR	InterPro:	IPR001590;	-
DR	InterPro:	IPR002870;	-
DR	PROSITE:	PS00142;	ZINC_PROTEASE; 1.
DR	PROSITE:	PS50215;	ADAM_MEPRO; 1.
DR	PROSITE:	PS50092;	TSP1; 2.
DR	PROSITE:	PS00427;	DISINTEGRINS; FALSE NEG.
KW	Hydrolase:	Metalloprotease;	Zinc; Signal; Glycoprotein; Zymogen;
KW	Repeat:	Extracellular matrix;	Heparin-binding.
FT	SIGNAL	1	POTENTIAL.
FT	PROPEP	29	BY SIMILARITY.
FT	CHAIN	229	ADAM-TS 8.
FT	METAL	378	ZINC (CATALYTIC) (BY SIMILARITY).
FT	ACT_SITE	379	BY SIMILARITY.
FT	METAL	382	ZINC (CATALYTIC) (BY SIMILARITY).
FT	METAL	388	ZINC (CATALYTIC) (BY SIMILARITY).
FT	DOMAIN	453	DISINTEGRIN-LIKE.
FT	DOMAIN	542	TSP-TYPE 1 1.
FT	DOMAIN	599	CYS-RICH.
FT	DOMAIN	706	SPACER.
FT	DOMAIN	847	TSP-TYPE 1 2.
FT	CARBOHYD	848	N-LINKED (GLCNAC. . . ) (POTENTIAL).
FT	CARBOHYD	415	N-LINKED (GLCNAC. . . ) (POTENTIAL).
FT	CARBOHYD	480	N-LINKED (GLCNAC. . . ) (POTENTIAL).
FT	CARBOHYD	506	N-LINKED (GLCNAC. . . ) (POTENTIAL).
FT	CARBOHYD	615	N-LINKED (GLCNAC. . . ) (POTENTIAL).
SO	SEQUENCE	905 AA;	124dd13b33a0cae CRC64;

[illegible]



Query Match	Best Local Similarity	Score	DB 1	Length	2569;
Matches	126;	Conservative	66;	Mismatches	218;
				Indels	221;
				Gaps	33;
QY 101	PAGSGGVYHGRP	REGFW	CLNREORPQNCN	132	
DB 503	PAGTGGH-HNCPEEGSALAGPMSSGASAAVROGYGPRYCKPCGRR	-----LCE	555		
QY 133	YVRFELCPGSLRRPRTERI	WSPAPMSKCSAAGOTGVOTRTICLAEMSLCS	186		
DB 556	YTGKCLCPRHRYRPOCEVEMNSFNHYPAGCDVCN-CSRGT	-----IEAVSEED	606		
QY 187	EASEEGQ-HCMGQDCTAC	DLTCPMGVNAD	CD	ACMCOF	224
DB 607	R-DSGQCRKPRVYGGQCDKCAPGLLPVLCVPCSCNRDGETSPSCDPTGACMKCE	-----	663		
QY 225	MHGAVALPGGAPASGAATYLLTKTPKLLTQDSDGRFVLPGLCPGKSLTKTKYKA	-----	283		
DB 664	---NVEGPOCLCREGSEFYLDPTNPKGCTC	---FCF-GVNEDCOSSRK-ORAKFVD	712		
QY 284	-----PIYLTMPKTSLSKA	-----	A	297	
DB 713	MNGWFPWRTADVDPPVSFNPSCNSKNVADLOELPVSYHSAWVAPRSTYGDKYVSTGYL	-----	772		
QY 298	TIKAEFVARETPYVMWNPETTKARAGQVSYLSCKATGKPRDPKYWHYNDTLTLLPSLYKH	-----	357		
DB 773	TYHAASFGLPDMVLLGKQPDVQLGQHMSLHKHESDRPRRL-HNRQVYIEGNEFH	-----	830		
QY 358	ESK-----LVLRKIQQ-HQAGEYFCKAQSDAGAVKSYADLLVTASDETPCNVP	-----	406		
DB 831	ESSSAPVSRREELMTVLPRLERLHIGLHPTETO	-----RLTLGVEGLEEASDGTGSG-P	882		
QY 407	ESYLRL-----PHDCONATNSPYVD	---YGRVCYKTCAGQDQDGCRAVQNC	454		
DB 883	RAHVECMACPPDYTGDSQCGCRPGYVWNKSLPVGRVCPCNCGNSN	---RCQDSGIC	939		
QY 455	--CGISKTEEREIQCSTGYLLPTKVAKEGSCQACTETRSIVRGVSAADNG	-----EPAR	506		
DB 940	INCOHNTAGEHEHROAGHYGNAIHGSCVYCCPHNNSATG-CAYDGAARACKP	-----	995		
QY 507	FGHYVMGNSRVSMTGKGTFTLHVPODTERLVLTVDRLQKFTVNTTKVLPENKKSAYEH	-----	566		
DB 996	-----GYTG	---QCERCAPGYEFGNQKFGGSCQPCNCSNG	1029		
QY 567	EIKMLRKREPTLLEMETNIIPLGEVYGGDP	597			
DB 1030	---QIGPCDPLT	---GDCVNDP	1046		

RESULT	8			
ID	TSPL_HUMAN	STANDARD:	PRT:	1170 AA.
AC	P07996			
DT	01-AUG-1988 (Rel. 08, Created)			
DT	01-AUG-1988 (Rel. 08, Last sequence update)			
DT	01-OCT-2000 (Rel. 40, Last annotation update)			
DE	THROMBOSPONDIN 1 PRECURSOR.			
GN	THBS1 OR TSPL OR TSP.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID:9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE-Endothelial cells;			
RX	MEDLINE-87057617; PubMed-2430973;			
RA	Lawler J., Hynes R.O.;			
RT	"The structure of human thrombospondin, an adhesive glycoprotein with			
RT	multiple calcium-binding sites and homologues with several different			
RL	proteins.";			
RL	J. Cell Biol. 103:1635-1648(1986).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE-89139590; PubMed-2918029;			
RA	Hennessey S.W., Frazier B.A., Kim D.D., Deckwerth T.L.,			
RA	Baumgartel D.M., Rotwein P., Frazier W.A.;			
RT	"Complete thrombospondin mRNA sequence includes potential regulatory			
RT	sites in the 3' untranslated region.";			
RL	J. Cell Biol. 108:729-736(1989).			
RN	[3]			
RP	SEQUENCE OF 1-397 FROM N.A.			
RX	MEDLINE-87157592; PubMed-3030396;			
RA	Kobayashi S., Eden-Mcutchan F., Framson P., Bornstein P.;			
RT	"Partial amino acid sequence of human thrombospondin as determined by			
RT	analysis of cDNA clones: homology to malarial circumsporozoite			
RT	proteins.";			
RL	Biochemistry 25:8418-8425(1986).			
RN	[4]			
RP	SEQUENCE OF 1-374 FROM N.A.			
RX	MEDLINE-86287276; PubMed-3461443;			
RA	Dixit V.M., Hennessey S.W., Grant G.A., Rotwein P., Frazier W.A.;			
RT	"Characterization of a cDNA encoding the heparin and collagen binding			
RT	domains of human thrombospondin.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 83:5449-5453(1986).			
RN	[5]			
RP	SEQUENCE OF 1-166 FROM N.A.			
RX	MEDLINE-89291870; PubMed-2544587;			
RA	Laherty C.D., Giernan T.M., Dixit V.M.;			
RT	"Characterization of the promoter region of the human thrombospondin			
RT	gene. DNA sequences within the first intron increase transcription.";			
RL	J. Biol. Chem. 264:11222-11227(1989).			
RN	[6]			
RP	SEQUENCE OF 1028-1170 FROM N.A.			
RA	la Fleur M., Jobin C., Gauthier J., Kreis C.G.;			
RL	Submitted (XXX-1992) to the EMBL/GenBank/DBJ databases.			
CC	-1- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND			
CC	CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN,			
CC	LAMININ AND TYPE V COLLAGEN.			
CC	-1- SUBUNIT: HOMOTRIMER, CROSS-LINKED BY DISULFIDE BONDS.			
CC	-1- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.			
CC	-1- SIMILARITY: CONTAINS 1 WFC DOMAIN.			
CC	-1- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.			
CC	-1- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.			
CC	-1- SIMILARITY: CONTAINS 7 TSP TYPE-3 DOMAINS.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			





```

RC TISSUE-COLON;
RX MEDLINE:91365376; PubMed-1679749;
RA Dodge G.R., Kovalesky I., Chu M.L., Hassell J.R., McBride O.W.,
RA Yi H.F., Iozzo R.V.;
RT "Heparan sulfate proteoglycan of human colon: partial molecular
RT cloning, cellular expression, and mapping of the gene (HSPG2) to the
RT short arm of human chromosome 1.";
RL Genomics 10:673-680(1991).
[4]
RN SEQUENCE OF 892-1398 FROM N.A.
RP TISSUE-Fibrosarcoma;
RX MEDLINE:92120660; PubMed-1685141;
RA Kallunki P., Eddy R.L., Byers M.G., Kestila M., Shows T.B.,
RA Tytgavason K.;
RT "Cloning of human heparan sulfate proteoglycan core protein,
RT assignment of the gene (HSPG2) to 1p36.1--p35 and identification of
RT a BamHI restriction fragment length polymorphism.";
RL Genomics 11:389-396(1991).
[5]
RN SEQUENCE OF 1-21 FROM N.A.
RX MEDLINE:94052171; PubMed-8234307;
RA Cohen I.R., Graessel S., Mutoch A.D., Iozzo R.V.;
RT "Structural characterization of the complete human perlecan gene and
RT its promoter.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:10404-10408(1993).
CC -1- FUNCTION: THIS PROTEIN IS AN INTEGRAL COMPONENT OF BASEMENT
CC MEMBRANES. IT IS RESPONSIBLE FOR THE FIXED NEGATIVE ELECTROSTATIC
CC CHARGE AND IS INVOLVED IN THE CHARGE-SELECTIVE ULTRAFILTRATION
CC PROPERTIES. IT INTERACTS WITH OTHER BASEMENT MEMBRANE COMPONENTS
CC SUCH AS LAMININ AND COLLAGEN TYPE IV AND SERVES AS AN ATTACHMENT
CC SUBSTRATE FOR CELLS.
CC -1- SUBUNIT: PURIFIED PERLECAN HAS A STRONG TENDENCY TO AGGREGATE IN
CC DIMERS OR STELLATE STRUCTURES.
CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC -1- TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANES.
CC -1- PTM: CONTAINS THREE HEPARAN SULFATE CHAINS AS WELL AS N-LINKED
CC AND O-LINKED OLIGOSACCHARIDES.
CC -1- SIMILARITY: CONTAINS 4 LDL-RECEPTOR CLASS A DOMAINS.
CC -1- SIMILARITY: CONTAINS 10.5 LAMININ EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 3 LAMININ DOMAINS IV.
CC -1- SIMILARITY: CONTAINS 22 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -1- SIMILARITY: CONTAINS 3 LAMININ G-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 4 EGF-LIKE DOMAINS.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X62515; CAA44373.1; -
DR EMBL; M85289; AAA52700.1; -
DR EMBL; M64283; AAA52699.1; -
DR EMBL; S76436; AAB21121.1; -
DR EMBL; L22078; NOT_ANNOTATED_CDS.
DR HSSP; P00740; IIXA.
DR MIM; 142461; -
DR InterPro: IPR000034; -
DR InterPro: IPR000082; -
DR InterPro: IPR000561; -
DR InterPro: IPR001438; -
DR InterPro: IPR001791; -
DR InterPro: IPR002049; -
DR InterPro: IPR002172; -
DR InterPro: IPR003006; -
DR Pfam: PF00008; EGF; 4.
DR Pfam; PF01390; SEA; 1.
DR Pfam; PF00047; Ig; 22.
DR Pfam; PF00052; laminin_B; 3.
DR Pfam; PF00053; laminin_EGF; 8.
DR Pfam; PF00054; laminin_G; 3.
DR Pfam; PF00057; Idl_recept_a; 4.
DR PRINTS; PR00010; EGFBL00D.
DR PROSITE; PS00022; EGF_1; 9.
DR PROSITE; PS01186; EGF_2; 5.
DR PROSITE; PS01209; LDLRA_1; 4.
DR PROSITE; PS01248; LAMININ_TYPE_EGF; 11.
DR PROSITE; PS50068; LDLRA_2; 4.
KW Signal: Basement membrane; Proteoglycan; Repeat; Glycoprotein;
KW Heparan sulfate; Laminin EGF-like domain; Immunoglobulin domain;
KW Extracellular matrix; EGF-like domain.
FT SIGNAL 1 21
FT CHAIN 22 4393
FT 22 193
FT 194 404
FT 405 506
FT 507 1678
FT 1679 3688
FT 3689 4393
FT 197 236
FT 283 321
FT 323 361
FT 366 405
FT 405 506
FT 523 532
FT 533 732
FT 733 765
FT 766 815
FT 816 873
FT 881 925
FT 926 935
FT 936 1127
FT 1128 1160
FT 1161 1210
FT 1211 1267
FT 1277 1326
FT 1327 1336
FT 1337 1531
FT 1532 1564
FT 1565 1614
FT 1615 1672
FT 1679 1773
FT 1774 1867
FT 1868 1957
FT 1958 2053
FT 2054 2153
FT 2154 2246
FT 2247 2342
FT 2343 2438
FT 2439 2535
FT 2536 2631
FT 2632 2728
FT 2729 2828
FT 2829 2926
FT 2927 3023
FT 3024 3114
FT 3115 3213
FT 3214 3300
FT 3301 3401
FT 3402 3490
FT 3491 3576
FT 3577 3671
FT 3671 3701
FT 3701 3847
FT 3846 3883
FT 3886 3924
FT 3966 4104
FT 4106 4143
FT 4145 4178
FT 4243 4391

```

```

FT SITE 65 67 HEPARAN SULFATE (POTENTIAL).
FT SITE 71 73 HEPARAN SULFATE (POTENTIAL).
FT SITE 76 78 HEPARAN SULFATE (POTENTIAL).
FT SITE 4151 4153 MEDIATES MOTOR NEURON ATTACHMENT
(POTENTIAL).
FT SITE 4301 4303 MEDIATES MOTOR NEURON ATTACHMENT
(POTENTIAL).
FT DISULFID 199 212 BY SIMILARITY.
FT DISULFID 206 225 BY SIMILARITY.
FT DISULFID 219 234 BY SIMILARITY.
FT DISULFID 285 297 BY SIMILARITY.
FT DISULFID 292 310 BY SIMILARITY.
FT DISULFID 304 319 BY SIMILARITY.
FT DISULFID 325 337 BY SIMILARITY.
FT DISULFID 332 350 BY SIMILARITY.
FT DISULFID 344 359 BY SIMILARITY.
FT DISULFID 368 381 BY SIMILARITY.
FT DISULFID 375 394 BY SIMILARITY.
FT DISULFID 388 403 BY SIMILARITY.
FT DISULFID 766 775 BY SIMILARITY.
FT DISULFID 768 782 BY SIMILARITY.

```

Query Match Best Local Similarity 2.4%; Score 151.5; DB 1; Length 4393;

Matches 149; Conservative 84; Mismatches 263; Indels 233; Gaps 36;

```

OY 84 RVCARPLRLARTTWTGAGTGVVHGSFREG-FWCINREORPCNGSNTRYFLCP 141
DB 3150 RISTPAKLEORTYGLMDSHYVLISSAKPSDAGTYVCL-AQNL- 3193
OY 142 GSIRRTERT-----WSPSPKSCSA-----CGOTGVQVIRT----- 174
DB 3194 GYAKQVEYIVDTGAMAPAPPOVAEAEIYVAGHTATLRCSATSPARTIMSKLRSP 3253
OY 175 -----RICALAEYSLCSSEASEGHCMDCTACDLTCEPGGVNA----- 214
DB 3254 LPMQHLEBDTLIRVAQODSGY-----ICNATSPAGHAETIILHVESPYATYV 3306
OY 215 -----DCDQACMCDPMLHG-----AVSLPGADASGAAYLLTKTKPLLT 254
DB 3307 PEHASVQAGETVLOQC-----LAHGTPLPTFOWMSRVSGSLPGRATARNELLHFERAP- 3359
OY 255 QVDSDEGRFRTPGLCPPGCKSLIKIT-KV-----KRAPIVLIMPKSLKAAITKAEVNAETP 309
DB 3360 --EDSGRYR-----CRVTNKGSAEAPQQLVGGPESLPTATSLPA-----GSP 3402
OY 310 YWVWNETARARAGOSVSLCKKATGKRPDKYFWYHNDLLDPSLKHHSKYLRLKLOOH 369
DB 3403 TVQVTPQLTKTSIGASVEFHC-AVPSDRGTQLRMFKEGQLPRGHSVQDGLRIQNLDS 3461
OY 370 QAGEYFCKAQSDAGAVKSKVAQLIVTASDETCNPYESTYLRLPHDCFONATNSFYDV 429
DB 3462 CGGTTCQAHHGPKAQAS-AQLVIA-----LP-SVLINI----- 3495
OY 430 GRPVTATGAGQDNGIRCDAYVONCGISKTERETQCGSYTLPTVAKECSQRTERR 489
DB 3496 -----RTSVQTVV-VGHAVEECLALGDKP-----QVTW 3524
OY 490 SIYRG--RVSADNGEPMFEGHYM--GNSRYSMTGKGTFTLHVPODTERLVLFVVR 544
DB 3525 SKVGHRLRGIVYSGGVVRIAHVELADACGYRTATNAGT-----TQSHVLLVQA 3576
OY 545 LQEVNTVTVLPNNKGSVAHFHEIKMLRKREPTTLEAMETNIIPLGEVNGE-DPMALF- 602
DB 3577 LPO-ISMPEVVRPAGSAAVE-----PCIASGYPTLPDISWSKLDGSLPPDRLEN 3625
OY 603 -----ISRSFYRNGNEPYI-----GKYKASTFLDPRNISTATAQIDPLNINDEGDT 651
DB 3626 NMLMLP--SVQPODAGTYVCTATNRGKVKAFANHLOVPERV-VPYPTQTPYSEFL----- 3676
OY 652 FLPLRTGMSVDFRDEVTSEPLNAGVKVYKHLSDTQVKMEHISTVYKMSLNP---TGLM 708
DB 3677 -PLPTIKDAVYRKFEIKITRPRDSADGMLX--NGQKRVG--SPYMLANRQPPFISFGLV 3731

```

```

OY 709 EEEGDKFE 717
DB 3732 GGRPERFD 3740

RESULT 10
NRG.DROME STANDARD; PRT: 1239 AA.
ID P20241; Q24414.
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE NEUROGLITAN PRECURSOR.
GN NRG.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta.
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 24-41 AND 737-751.
RX MEDLINE=90030418; PubMed=2805067;
RA Bieber A.J., Snow P.M., Hortsch M., Patel N.H., Jacobs J.R.,
RA Traquina Z.R., Schilling J., Goodman C.S.;
RT "Drosophila neuroglian: a member of the immunoglobulin superfamily
RT with extensive homology to the vertebrate neural adhesion molecule
RT L1."
RL Cell 59:447-460(1989).
RN [2]
RP SEQUENCE OF 1182-1239 FROM N.A.
RX MEDLINE=90262720; PubMed=1693086;
RA Hortsch M., Bieber A.J., Patel N.H., Goodman C.S.;
RT "Differential splicing generates a nervous system-specific form of
RT Drosophila neuroglian."
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 610-814.
RX MEDLINE=94213741; PubMed=7512815;
RA Huber A.H., Wang Y.-M.E., Bieber A.J., Bjorkman P.J.;
RT "Crystal structure of tandem type III fibronectin domains from
RT Drosophila neuroglian at 2.0 A."
RL Neuron 12:717-731(1994).
CC -1- FUNCTION: THIS PROTEIN MAY PLAY A ROLE IN NEURAL AND GLIAL CELL
CC ADHESION IN THE DEVELOPING DROSOPHILA EMBryo.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A LONG FORM (SHOWN HERE) AND A
CC SHORT FORM. ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: NEURONS AND GLIA IN THE DEVELOPING NERVOUS
CC SYSTEM AND ON SOME OTHER NONNEURONAL TISSUES.
CC -1- SIMILARITY: CONTAINS 6 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -1- SIMILARITY: CONTAINS 5 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib.ch).
CC -----
DR EMBL: M28231; AAA28728.1; ALT-SEQ.
DR EMBL: X76243; CAAS3822.1; -.
DR PIR: A32579; A32579.
DR PDB: 1CFB; 30-NOV-94.
DR FlyBase: FBgn0002968; Nrg.
DR InterPro: IPR001777; -.
DR InterPro: IPR003006; -.
DR Pfam: PF00041; fn3; 5.
DR Pfam: PF00047; lg; 6.
DR Cell adhesion: Glycoprotein; Transmembrane; Repeat; 3D-structure;
KW Immunoglobulin domain; Signal; Embryo; Alternative splicing.
FT SIGNAL 1 23

```

```

FT CHAIN 24 1239 NEUROGLIAN.
FT DOMAIN 24 1138 EXTRACELLULAR (POTENTIAL).
FT TRANSLEM 1139 1154 POTENTIAL.
FT DOMAIN 1155 1239 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 53 123 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 149 224 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 262 329 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 354 422 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 447 512 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 536 606 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 629 690 FIBRONECTIN TYPE-III.
FT DOMAIN 729 792 FIBRONECTIN TYPE-III.
FT DOMAIN 832 896 FIBRONECTIN TYPE-III.
FT DOMAIN 932 997 FIBRONECTIN TYPE-III.
FT DOMAIN 1024 1098 FIBRONECTIN TYPE-III.
FT DISULFID 59 111 POTENTIAL.
FT DISULFID 625 706 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 182 182 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 198 198 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 411 411 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 652 652 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 683 683 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 821 821 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1125 1125 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 1234 1234 T -> Y (IN REF. 2).
FT CONFLICT 1237 1237 L -> K (IN REF. 2).
SQ SEQUENCE 1239 AA; 138284 MM; 49EL2692DDDD027D CRC64;

```

Query Match Best Local Similarity 2.3%; Score 146; DB 1; Length 1239; Matches 149; Conservative 99; Mismatches 277; Indels 208; Gaps 41;

```

QY 304 VRAETPYVMNPEFKARRAGQSVSLCKATGPRPRKYFWYHNDTLDDPSLYKHESK--L 361
D 427 VQAPPTISEAPAVSVYDGRNVTIKRCVNGSPKP-LVKMLRASWMLGGRYVNGDGL 485
QY 362 VLRLQOHQAGEYFCKKASDAGAVKVAQLIVASDETPCNPPESTLILPHDCFQNA 421
D 486 EIDQVTSADACKYCYAONKPFGEIAD-GSLIV--KEHTRITOEPONEYVA---AGQSA 538
QY 422 T---NSFYDVGRCPRVKCA--GO-----ODNGIRCRBAVONCGISKEEREIOCSGYT 471
D 539 TFRONEAHDLLEIEMWMDGQSIDPEAQPRFKYND--NSLTIAIT--MELDSGEYT 593
QY 472 LPTVARECSQCRCTEERSIVR-----GRVSAADNGEPMRGHYMGNRSYMTGY 522
D 594 C---VARTRLDEATVARNLIVQDVNAPKLTGICQADKAE---IHWEQGDNSPILHY 647
QY 523 KGTG-TLHVP-----QDTERLVLTFFYDLQKFEVN--TTKVLPRNKKGSA--VHEIKML 571
D 648 TIOENTSTFPASMDAAVEKVENPTDSSFVQMSPMANYTFRIAYFNKIGASPPSAHSDSCT 707
QY 572 R-----KEPTLEAMETNIIPLGEVGEDMAELE-----IPSRNF--- 608
D 708 TQDPVPEKNPNVVGQGTG--PNNLVISWTMPLEIENHAPNFHYVSKRQIPAAWENN 765
QY 609 ---YRONG---EPYIGKVASVTFDPNINISTATTAQTDLNFINDGTFPLRTYGM 659
D 766 NIFPMRQNNIYIADQPTFVKLKIVAINDRGESVAAEEV---VQSGEDRPLDAPTIN 821
QY 660 FSVDFRDEVTSE---PLNAGKAVVHLDSIOVKMPEHISTYVKLMSLNDPTGLMEEE 711
D 822 FTM--ROITSTSGYAMTPIVSESVGHFKGYK-----IQWTW-----FNE 861
QY 712 GDFKFNOR--RNKREDTFLVGNLEIRERLJFNLDVPSRRCFKAAYASERFLPSEQ 769
D 862 GE---EGIRELHVKGDTHNALVTOFK-----PDSKN-YAKILAIYKCR----- 899
QY 770 IQGVIVISINLEPRTGLFSLNPRAMGRFDSVITGPNGACVPAFCDDQSDPAY---SAYVL 825
D 900 -----FNGPPSAVIDFDT-----PEGVSPV-----QGLDAVPLIGSSATML 935

```

```

QY 826 -----ASTAGEELOAVESPSPKFNMAIGVPOPLYLNRYRRTDHEPRVKKTAFOI 876
D 936 HMKRPLPNGLTGKTYEE-----VASEYGEKREKDPHTDPRVTR-----M 980
QY 877 SMARPPNS-----AEESNGPIYAFE--NRACEAPPSAHPFYQIECD----R 921
D 981 KMGKLPKSKRYSITATATKKGESSEHYIEKTLKDAVNVAVPATPSSWEQLPSDNGLAK 1040
QY 922 YDVTVPFNEDDP 934
D 1041 FRIMWLPSTEGHP 1053

RESULT 11
TSPL_XENLA
ID TSPL_XENLA STANDARD; PRT; 1173 AA.
AC P35448;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE THROMBOSPONDIN 1 PRECURSOR.
GN THBS1 OR TSPL.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RA urry L.A., Ramos J., Duquette M., Deslone D.W., Lavier J.;
RL Submitted (xxx-1993) to the EMBL/GenBank/DBJ databases
CC -1- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND
CC CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN,
CC LAMININ AND TYPE V COLLAGEN.
CC -1- SUBUNIT: HOMOTRIMER, CROSS-LINKED BY DISULFIDE BONDS.
CC -1- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.
CC -1- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
CC -1- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.
CC -1- SIMILARITY: CONTAINS 7 TSP TYPE-3 DOMAINS.
CC -----
CC This SWISS-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L04278; -; NOT_ANNOTATED_CDS.
DR HSSP; P00740; IIXA.
DR InterPro; IPR000561; -.
DR InterPro; IPR000884; -.
DR InterPro; IPR001007; -.
DR Pfam; PF00008; EGF; 2.
DR Pfam; PF00090; TSP_1; 3.
DR Pfam; PF00093; vwc; 1.
DR PROSITE; PS00022; EGF_1; FALSE_NEG.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS00092; TSPL; 3.
DR PROSITE; PS01208; vwc; 1.
DR Glycoprotein; Cell adhesion; Calcium-binding; Heparin-binding; Repeat;
KW EGF-like domain; Signal.
FT SIGNAL 1 22
FT CHAIN 23 1173 THROMBOSPONDIN 1.
FT DOMAIN 23 235 HEPARIN-BINDING (POTENTIAL).
FT DOMAIN 319 376 VWFC.
FT DOMAIN 382 433 TSP TYPE-1 1.
FT DOMAIN 438 494 TSP TYPE-1 2.
FT DOMAIN 495 546 TSP TYPE-1 3.
FT DOMAIN 550 590 EGF-LIKE 1.
FT DOMAIN 591 648 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).

```



```

DR EMBL: M96655; AAA49083.1;
DR EMBL: M88283; AAA8647.1;
DR EMBL: M88284; AAB53768.1;
DR PIR: A25810; A25810.
DR PIR: A37099; A37099.
DR PIR: S11652; S11652.
DR InterPro: IPR000719;
DR InterPro: IPR001777;
DR InterPro: IPR002290;
DR Pfam: PF00041; fn3; 1.
DR Pfam: PF00047; Ig; 9.
DR Pfam: PF00069; Pkinase; 1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS00108; PROTEIN_KINASE_SF; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
DR TransErase; Serine/threonine-protein kinase; Calmodulin-binding;
KW ATP-binding; Phosphorylation; Immunoglobulin domain; Repeat;
KW Alternative initiation.
FT CHAIN 1 1906
FT CHAIN 935 1906
FT CHAIN 1750 1906
FT INIT_MET 935 1906
FT INIT_MET 1750 1906
FT DOMAIN 1330 1400
FT DOMAIN 1453 1708
FT DOMAIN 1809 1876
FT NP_BIND 1459 1467
FT BINDING 1482 1482
FT ACCT_SITE 1574 1574
FT DOMAIN 1716 1728
FT DOMAIN 1730 1749
FT DOMAIN 1317 1364
FT DOMAIN 1385 1402
FT DOMAIN 660 1833
FT REPEAT 660 676
FT REPEAT 758 774
FT REPEAT 1107 1123
FT REPEAT 1817 1833
FT DOMAIN 693 1866
FT REPEAT 791 807
FT REPEAT 1140 1156
FT REPEAT 1281 1297
FT REPEAT 1851 1866
FT DOMAIN 970 1226
FT REPEAT 970 987
FT REPEAT 999 1016
FT REPEAT 1061 1078
FT REPEAT 1209 1226
FT DOMAIN 1700 1763
FT DOMAIN 1896 1906
FT MOD_RES 1748 1748
FT MOD_RES 1762 1762
FT CONFLICT 1439 1439
SQ SEQUENCE 1906 AA; 210445 MW; AD7DBAJB69EE363 CRC64;

Query Match 2.3% Score 145; DR 1; Length 1906;
Best Local Similarity 19.3%; Pred. No. 0.074;
Matches 179; Conservative 109; Mismatches 305; Indels 334; Gaps 45;
QY 140 PGSLRPTERISWSPWSKCSAAGOTGTRICIAEAVSLCSASEE-----GQ 193
DB 1111 PPAVY-----SWTLDSK-----AIKSSKSIYISQEGTLCSLTIKVMEDDGE 1153
QY 194 H-CMGQDTACDLTCGPMQVNAADCAKQCPMLHGAWSLFGGAPASCAAIYLLTKTKRL 252
DB 1154 XKCAENMAG-----KAEC-AC---KVLVEDTSTSTRAAKPAEKRTKPKTLPV 1199

```

```

QY 253 LTQTDSDGRFRIPGLCPDQKSLITTKYKPAPIVLTMPKTSIAKATIKAEVRAETPYMV 312
DB 1200 LTESSE-----ATVKKRP-----AKTPPKAAT-----PQIT 1228
QY 313 MNPETKARAGOSVSLCCATGKPRPDXYFWYHNDTLDPSLY-----KHESKLVLRKL 366
DB 1229 QPEDRKVRAGESVLEFPAKVVG--TAPITCYMKFRKQIQEENYIKIENAEKSLTIST 1287
QY 367 QOHQAGEYFCKAQSDAGAKSKVAQLIVTASDE--TPC-NPVPSYLIRLPHDCFQONAT 422
DB 1288 KOEHGCGYTLVVENKLSROAQVNLTVDKDPDPAGTPCASDISSSSL-----T 1336
QY 423 NSFY--DVGRC-----PYKTCAG-----QOONGIRCRPAVONC 454
DB 1337 LSWGSSYDGSANOSYVEIWNVSVDNKMWTDTLTCTRSFNVQDLQADREYKFRRAANV 1396
QY 455 CGIS-----KTEEREIQSGYTLPTKYAKESQCRCTETRSIVRGVSAADNG 502
DB 1397 YGISPPQSESEVYKKGQEEBELKEEAEALSDDECKETEVNRYRTITN--TEOKVSDVYNI 1455
QY 503 EPM---RFGHYMGNSRVSMTGYKGT-----TLHPQDTERLV 538
DB 1456 EERLGSFGGVFLVERKTKGVWAGKFKFAYSAKEKENIRDEISIMNCLHHPK-----L 1510
QY 539 LTFVDRLOKFNVTFTVLPFNKKGSAVFHEIKMLRKKEPTITLEAMETNITPLGEVVGEDPM 598
DB 1511 VQCVDAFEKANIYVWLDLW-VSGGELFERI-----IDED-- 1543
QY 599 AELEIPSR---SFYRO--NGEPYIGKRVASVTFDPNINISTATAQTDLPINDEGDFRP 653
DB 1544 --FELTERECIKYMQQISEGVEYIHKQGIYVHLDLKPENIMCVNKTGTISIKLID----- 1594
QY 654 LRTYGMSEVDFREYTSPLNAGKVKVHLDSYQVMPHIS-----TVKLMS----- 700
DB 1595 -----FOLARLE-----SAGSLKLVFGTPEFAVPEVINEPIGYETDWMISGIVCIYI 1642
QY 701 ---LNPDTGLMEER-----GDFKFNORRNK--REORTFLVGNLLEIRERFLVDPE 748
DB 1643 LVSGSLPMDGNDNETLANVTSATWDFDEADPELSDAKDPIS-----LKKDKMS 1695
QY 749 SRRC-----FVKVAYRSERFLPSEIQGVVISVNLNERTGFLSNPR 791
DB 1696 RLNCQCCQHWLQKDTNMKAKKLSKDKMKYMARRWQ-----KTGAA--VR 1742
QY 792 AMGRFDS--VITPGNACVPAFCDDQSPDAYSAVYLAELQAVSSPKFNPAIGV 849
DB 1743 AIGRLSNAMISGMSGR-----KAGSSP----- 1766
QY 850 PQPLYLNKINRYRTHEDRVRKKTAFQISMARPRPSAESNGPIYAFENLRACEAPPSA 909
DB 1767 ---TSPLNACKVENEDAFLEVAEKPRVPR-----YFTKTIIDMEVEGSA 1810
QY 910 AHFRFYQTEGDYR-DYNTVPFNEEDPM 935
DB 1811 ARFD-CKIEG--YPDPEVMWYKDDQPV 1834

RESULT 13
KMLS_BOVIN STANDARD; PRT; 1176 AA.
AC 028824;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE MYOSIN LIGHT CHAIN KINASE, SMOOTH MUSCLE (EC 2.7.1.117) (MLCK)
DE [CONFAINS: TELOKIN].
GN MYLK.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
NCBI_TaxID=9913;

```

FT	BINDING	754	754	ATP (BY SIMILARITY).
FT	ACT_SITE	846	846	BT SIMILARITY.
FT	DNAIIN	1171	1176	POLY-GUO.
SO	SEQUENCE	1176 AA; 128824 MW; F53DC6D4D42D4B97 CRC64;		

  

47;

```

Query Match          2.3%; Score 144.5; DB 1; Length 1176;
Best local similarity 20.3%; Pred. No. 0.04; Mismatches 306; Indels 281; Gaps
Matches 179; Conservative 115;

QY 166 GGTGGVQTRFICLAEMWSLCSASEEHCOCMOD-----CTACDLTCGPMVNADCDACM 220
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db   393 GKT-LKTKRKFIVLSQEGSLCSVSTIEK---ALPEDRGIKCYAKN--SAGAASESCQV-- 443

QY 221 CODEPHLHVAVSLPGAPASGAAILYTLPRLYLQTDSGDGRFRIRPLGLCPDCKSLTKITKV 280
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db   444 -----TVDPV-DAPTSENA-----KAPMKARRPKSSLDPVLGIESD-----ATV 482

QY 281 KFAPIVLTMPRTSLKAATIKAEFAAETPRVVMPEETARAGOSVSLCATKGKPRDK 340
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db   483 KKKP----AKPTPPKAA-----PQLTIQRPEODKVAAGESVELFGVAG-TQPIIT 528

QY 341 YFWY-HNDTLLDSPLYKHE-----SKVLARKLOOHQAQEYFCKAQSADAGAVSKVAQDLIV 394
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db   529 CTWAKFKKOIDSEHIKENESOGSKLIIRAORHCQCYTLLVENKLGSROAQVNLTVV 588

QY 395 TASDE-TPC-NPYRESYLRLPHDCQNNTNFY---YDVGRCPVKYTCAGOODNGIRC 447
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db   589 DKPPRPACTPCASDIRSSSL-----TLMSYGSDGG-----SAVSYSVEI 630

QY 448 RDAVO-----NCCGIS-----KTEBREIOCGSYTLPTKYAKESCOCRET 488
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db   631 WDSVDKTMKELATCRHSFNQDLLPREYKFRVRALINVGTSPESQSELTLALGEKPEE 690

QY 489 RSIVRGKRSADNDNEPARFGHYVMGNNSVMTGKGFTHLVPODTERLVUTFPD---RL 545
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db   691 EPKDEVEVSDDEKEP-----EVDR-RVTYV---NTEOKVSDFYDIERL 731

QY 546 -QKRVNTTKVLPPNKSGSAV---FHBIKMRLRKREPTLEAM----- 582
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db   732 GSGKFGQGVFRIV-EKKTGKIWKAGFKKAYSAKAKENIROEISIMNCLHKFLVCYDAF 789

QY 583 --ETNIITPLGCVGEDPMAE-----LEIPSRFSYKONGEPYIGKVKASYTELDPNNIST 634
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db   790 EEKANIVAVLEIVSGGELFERIIDEDDELTERECIK-----YMKOISEGYEIHKQGIVH 844

QY 635 ATAAQTDLNFINDSGDREFPLTYGMFVSVDPRDEVTSPLNAGKVKYILDSTQVMMKHIS 694
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db   845 LDKLPENIMCYNKTGTRIKLDFGL-----ARRLENAGSLKVLGPFEVPYAPRYIN 895

QY 695 -----TVKIMS-----LNPDGLMEEE-----GDFKFFNRRNK-REDRTF 729
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db   896 YEPIGATWDMSISIVICYIIYVGLSPFMGDNDNETLANVTSATWDFDEADELSDDAKD 955

QY 730 LVGLLELRERRLFMLDVPESRRC-----FVKRAAIRSEFLPSIEDIQ 772
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db   956 FISH-----LLKKDMKNRLNCTQCLQHPLMKDTNMEEKKILSKDRMKYARRRWKO- 1007

QY 773 VVISVINLEPRTGFLSNPRAMGRFDVSITGPNGCACYPAFCDOSPDAYSAAYVLASLAGEE 832
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db   1008 -----KTG-NAVRAIIGRLSMA-----MISGLSGRK 1032

QY 833 LQAVESSPKERNPAIGVPPQIYLNKLANTRRIDHDHPRYKKTAFQIISMAKPPRNSAESNGP 892
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db   1033 -----SSTGSP--TSPLNAEKLESBD--VSQAFLFAVAVEEKPRPHVKP----- 1069

QY 893 IYAENLRACEAAPPSAAHFQYOIEDGR-DVYTFPENED 932
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db   1070 -YFSKTRIIDELEVYEGSARD-CRIEG-YDPPEVWFVKD 1106

RESULT 14
RMLS_HUMAN

```

ID KML5\_HUMAN STANDARD; PRT: 1914 AA.  
AC Q15746; Q14844; Q16794; Q95796; Q95797; Q95798; Q95799; Q9UBG5;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 01-OCT-2000 (Rel. 40, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE MYOSIN LIGHT CHAIN KINASE, SMOOTH MUSCLE AND NON-MUSCLE ISOZYMES  
DE (EC 2.7.1.117) (MCK) [CONTAINS: TELOKIN (KINASE RELATED PROTEIN)  
DE (KRP)].  
GN MYLK OR MLCK.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OC NCBI\_Taxid:9606;  
OX [1]  
RN SEQUENCE FROM N.A. (ISOFORM 1).  
RP TISSUE-Unbiblical vein endothelial cells;  
RX MEDLINE=97304466; Pubmed=9160829;  
RA Garcia J.G.N., Lazar V., Gilbert-Mcclain L.I., Gallagher P.J.,  
RA Verin A.D.;  
RT "Myosin light chain kinase in endothelium: molecular cloning and  
RT regulation.";  
RL Am. J. Respir. Cell Mol. Biol. 16:489-494(1997).  
RN [2]  
RP REVISIONS.  
RA Birukov K.G., Garcia J.G.N.;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A. (ISOFORMS 2; 3A; 3B AND 4).  
RX TISSUE-Unbiblical vein;  
MEDLINE=99216419; Pubmed=10198165;  
RA Lazar V., Garcia J.G.N.;  
RT "A single human myosin light chain kinase gene (MLCK; MYLK).";  
RL Genomics 57:256-267(1999).  
RN [4]  
RP REVISIONS TO ISOFORM 2.  
RA Birukov K.G., Garcia J.G.N.;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RP SEQUENCE OF 923-1914 FROM N.A.  
RX TISSUE-Hippocampus;  
MEDLINE=96121365; Pubmed=8575746;  
RA Potter M.-C., Cheiot E., Pekarsky Y., Gardiner K., Rossier J.,  
RA Turnell W.G.;  
RT "The human myosin light chain kinase (MLCK) from hippocampus:  
RT cloning, sequencing, expression, and localization to 3qgen-q21.";  
RL Genomics 29:562-570(1995).  
RN [6]  
RP SEQUENCE OF 1614-1914 FROM N.A.  
RX TISSUE-Lung, and Placenta;  
MEDLINE=20007838; Pubmed=10536370;  
RA Matterson D.M., Schavocky J.P., Guo L., Weiss C., Chlenski A.,  
RA Shrivastava V.P., Van Eldik L.J., Halech J.;  
RT "Analysis of the kinase-related protein gene found at human chromosome  
RT 3q21 in a multi-gene cluster: organization, expression, alternative  
RT splicing and polymorphic marker.";  
RL J. Cell. Biochem. 75:481-491(1999).  
RN [7]  
RP SEQUENCE OF 1456-1914 FROM N.A.  
RX TISSUE-Placenta;  
RA Matterson M.D.;  
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: CALCIUM/CALMODULIN-DEPENDENT ENZYME RESPONSIBLE FOR  
CC SMOOTH MUSCLE CONTRACTION VIA PHOSPHORYLATION OF A SPECIFIC SERINE  
CC IN THE N-TERMINUS OF MYOSIN LIGHT CHAINS (MLC), AN EVENT THAT  
CC FACILITATES MYOSIN INTERACTION WITH ACTIN FILAMENTS. CENTRAL  
CC DETERMINANT IN THE DEVELOPMENT OF VASCULAR PERMEABILITY AND TISSUE  
CC EDEMA FORMATION. IN THE NERVOUS SYSTEM IT HAS BEEN SHOWN TO  
CC CONTROL THE GROWTH INITIATION OF ASTROCYTIC PROCESSES IN CULTURE  
CC AND TO PARTICIPATE IN TRANSMITTER RELEASE AT SYNAPSES FORMED  
CC BETWEEN CULTURED SYMPATHETIC GANGLION CELLS. CRITICAL PARTICIPANT  
CC IN SIGNALING SEQUENCES THAT RESULT IN FIBROBLAST APOPTOSIS.  
CC -1- CATALYTIC ACTIVITY: ATP + [MYOSIN LIGHT-CHAIN] = ADP + [MYOSIN  
CC LIGHT-CHAIN] PHOSPHATE.

CC -1- SUBUNIT: TELOKIN BINDS CALMODULIN.  
CC -1- ALTERNATIVE PRODUCTS: AT LEAST 3 ISOFORMS OF MLCK ARE PRODUCED  
CC BY ALTERNATIVE INITIATION, A NON-MUSCLE FORM (WHICH IS THE LONGEST  
CC FORM); A SMOOTH-MUSCLE FORM AND TELOKIN (A C-TERMINAL SECTION WITH  
CC NO CATALYTIC ACTIVITY).  
CC -1- ALTERNATIVE PRODUCTS: AT LEAST 6 ISOFORMS; 1 (SHOWN HERE), 2, 3A,  
CC 3B, 4 AND DEL-1790; ARE PRODUCED BY ALTERNATIVE SPLICING.  
CC -1- TISSUE SPECIFICITY: SMOOTH MUSCLE AND NON-MUSCLE ISOZYMES ARE  
CC EXPRESSED IN A WIDE VARIETY OF ADULT AND FETAL TISSUES AND IN  
CC CULTURED ENDOTHELIUM WITH QUALITATIVE EXPRESSION APPEARING TO BE  
CC NEITHER TISSUE- NOR DEVELOPMENT-SPECIFIC. NON-MUSCLE ISOFORM 2 IS  
CC THE DOMINANT SPICE VARIANT EXPRESSED IN VARIOUS TISSUES. TELOKIN  
CC HAS BEEN FOUND IN A WIDE VARIETY OF ADULT AND FETAL TISSUES.  
CC -1- PFM: MLCK IS PROBABLY DOWN REGULATED BY PHOSPHORYLATION.  
CC -1- SIMILARITY: BELONGS TO THE SPR/THR FAMILY OF PROTEIN KINASES.  
CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.  
CC -1- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE III-LIKE DOMAIN.  
CC -----  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: U48959; AAC18423.2; -;  
DR EMBL: AF069601; AAD15921.2; -;  
DR EMBL: AF069602; AAD15922.1; -;  
DR EMBL: AF069603; AAD15923.1; -;  
DR EMBL: AF069604; AAD15924.1; -;  
DR EMBL: X85337; CAA59685.1; -;  
DR EMBL: AF096771; AAD51380.1; -;  
DR EMBL: AF096766; AAD51380.1; JOINED.  
DR EMBL: AF096767; AAD51380.1; JOINED.  
DR EMBL: AF096768; AAD51380.1; JOINED.  
DR EMBL: AF096769; AAD51380.1; JOINED.  
DR EMBL: AF096770; AAD51380.1; JOINED.  
DR EMBL: AF096771; AAD51381.1; -;  
DR EMBL: AF096769; AAD51381.1; JOINED.  
DR EMBL: AF096770; AAD51381.1; JOINED.  
DR EMBL: X90870; CAA62378.1; -;  
DR HSSP: P56276; IRLK.  
DR MIM: 600922; -;  
DR InterPro: IPR000719; -;  
DR InterPro: IPR001777; -;  
DR InterPro: IPR002290; -;  
DR InterPro: IPR003006; -;  
DR Pfam: PF00041; fn3; 1.  
DR Pfam: PF00047; fg; 8.  
DR Pfam: PF00069; pkinase; 1.  
DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE: PS00108; PROTEIN\_KINASE\_ST; 1.  
DR PROSITE: PS50011; PROTEIN\_KINASE\_DOM; 1.  
KW Transferrase: Serine/threonine-protein kinase; Calmodulin-binding;  
KW ATP-binding; Phosphorylation; Immunoglobulin domain; Repeat;  
KW Alternative initiation; Alternative splicing.  
FT CHAIN 1 1914  
FT MYOSIN LIGHT CHAIN KINASE, NON-MUSCLE  
FT ISOZYME.  
FT CHAIN 923 1914  
FT MYOSIN LIGHT CHAIN KINASE, SMOOTH-MUSCLE  
FT ISOZYME.  
FT CHAIN 1761 1914  
FT TELOKIN.  
FT INT\_MET 923 923  
FT FOR MYOSIN LIGHT CHAIN KINASE, SMOOTH-  
FT MUSCLE ISOZYME.  
FT INT\_MET 1761 1761  
FT FOR TELOKIN.  
FT DOMAIN 1343 1413  
FT FIBRONECTIN TYPE-III.  
FT DOMAIN 1464 1719  
FT PROTEIN KINASE.  
FT DOMAIN 1711 1774  
FT CALMODULIN-BINDING.  
FT DOMAIN 1824 1891  
FT IG-LIKE C2-TYPE DOMAIN.  
FT NP\_BIND 1470 1478  
FT ATP (BY SIMILARITY).  
FT BINDING 1493 1493  
FT ATP (BY SIMILARITY).  
FT ACT\_SITE 1585 1585  
FT BY SIMILARITY.

```
FT DOMAIN 1906 1914 POLY-GLU.
FT DOMAIN 868 998 5 X 28 AA APPROXIMATE TANDEM REPEATS.
FT REPEAT 868 895 1-1.
FT REPEAT 896 923 1-2.
FT REPEAT 924 951 1-3.
FT REPEAT 952 979 1-4.
FT REPEAT 980 998 1-5 (INCOMPLETE).
FT DOMAIN 999 1063 6 X 12 AA APPROXIMATE TANDEM REPEATS.
FT REPEAT 999 1003 11-1 (INCOMPLETE).
FT REPEAT 1004 1015 11-2.
FT REPEAT 1016 1027 11-3.
FT REPEAT 1028 1039 11-4.
FT REPEAT 1040 1051 11-5.
FT REPEAT 1052 1063 11-6.
FT VARSPLIC 437 506 VSGIPREYAMFLEGEPVAPROGSEIEVEDAGSHYLCILKA
FT RTRDSGTSCASNAGOVSCMTLOVER -> G (IN
FT ISOFORM 2 AND ISOFORM 3B).
FT VARSPLIC 1433 1439 DEVEYSD -> MKMRCOT (IN ISOFORM 3A,
FT VARSPLIC 1473 1546 ISOFORM 3B AND ISOFORM 4).
FT GFGGVFLVEKTRKVMAGKFKFSAKEKENIROELISIM
FT NCLHHPKLVOCYDAFEKEKANIWMLEIVSGGEL -> L
FT (IN ISOFORM 4).
FT VARSPLIC 1655 1705 MISSING (IN ISOFORM 3A AND ISOFORM 3B).
FT VARSPLIC 1790 1790 MISSING (IN ISOFORM DEL-1790).
FT V -> M (IN REF. 5).
FT 963 963 S -> P (IN AAD15922).
FT 1022 1022 P -> A (IN REF. 5).
FT 1042 1042 KPM -> EAH (IN REF. 5).
FT 1162 1162 P -> L (IN AAD15922 AND AAD15923).
FT 1210 1210 L -> P (IN REF. 5).
FT 1280 1280 E -> D (IN AAD15922 AND AAD15923).
FT 1284 1284 M -> I (IN AAD15922, AAD15923 AND
FT AAD15924).
FT 1300 1300 A -> G (IN REF. 5).
FT 1316 1316 L -> S (IN REF. 5).
FT 1326 1326 T -> S (IN REF. 5).
FT 1478 1478 V -> C (IN REF. 5).
FT 1511 1511 S -> T (IN AAD15922 AND AAD15923).
FT 1563 1563 I -> T (IN REF. 5).
FT 1609 1609 A -> P (IN REF. 5).
FT 1639 1639 G -> R (IN REF. 5).
FT 1640 1640 GY -> D (IN AAD15922, AAD15923 AND
FT AAD15924).
FT 1648 1648 G -> R (IN REF. 5).
FT 1659 1659 LS -> P (IN REF. 5).
FT 1710 1711 CT -> LA (IN REF. 5).
FT 1897 1897 L -> H (IN AAD15922, AAD15923 AND
FT AAD15924).
SQ SEQUENCE 1914 AA: 210773 MW: 9295FE0279CB0033 CRC64;
```

Query Match 2.3%; Score 144.5; DB 1; Length 1914;  
Best Local Similarity 20.7%; Pred. No. 0.08;  
Matches 221; Conservative 133; Mismatches 357; Indels 357; Gaps 58;

```
Db 1184 QVTV-----DDAPASEN-----TKAPEMKSRRPKSSLPVLGTESD----- 1219
Qy 277 ITVKKFAPVILVMPKTSKATIKAEFVRAETPYMNPETKARRAGOSVSLCCATKGRP 336
Db 1220 -ATVKKRP--APKTPPRKAM-----PQIITQFEDQVVRAGESVELFGKVTG-T 1264
Qy 337 RPDKRFWYHNDTLLEDLSL-----KHESKLVLRKLOHQAGEYFCKAOSDAGAVASKVA 390
Db 1265 QPITCTWKKFRKQIQSEHMKVSENSKSLTILARQEHCGCYTLLENKIGSRQAQVN 1324
Qy 391 QILIVTASDE---TPC-----NPVESYLRLPHCFONATNSFY 427
Db 1325 LTVVADPDPPAGTPCASDIRSSLSLTLSWYSSYDGAASQVSYIEL-----WDSANKTKM 1379
Qy 428 DVGRCPVKTCAGQQ-----DNGIRCDAVONOCGISK--TEEREICSCGYTLPTVAKKCSO 483
Db 1380 ELATCSTSNVDDLPDHEIKFRVAINVYGSFSGQSESLTTVG-----E 1426
Qy 484 RCTETRSIVGRVSAADNCEPMREGHVYMGNSRVMTGYKGFTHVPODTERLVLFVND 543
Db 1427 KPEPRDEV--EVSDDDEKEP-----EVDYR-TVYI--NTBQKVSDFYD 1465
Qy 544 ---RL--QKFTVNTKVLPRKKSAY-----FHEIKMLRKEPTITLAM----- 582
Db 1466 IEERLSGGFGQVFRVY--EKTRKWAGKFKAYSAKEKENIROELISIMNCLHHPKLVQ 1523
Qy 583 -----ETNIPLGHEVGDPMAB-----LEIPSR--SFYRQ--NGEPYIGKAKSV 624
Db 1524 CVDAFEKANIWVLEIVSGGELFERLIDEDFELTRECKIKYRQISEGVEYIHKQGIYH 1583
Qy 625 TFLDPNNISTATAQTDLNFINDGTFPLRTYGMFSVDFRDEVTSEPLNACKVKVHLD 684
Db 1584 LDKPENIMKVNKGTIRKILID-----FGLARLE-----NAGSLKVLFGT 1624
Qy 685 TOYKMEPHIS-----TVKLMS-----LNPDTGIMEE-----GOFKRENR 720
Db 1625 PEFVAPEVNIYEPYIGATDWMISGYICILYVSGLSPEPMGNDNETLANVTSATWDDDEE 1684
Qy 721 RNK-REDRTFLVGNLEIRERKRLFNLDVPESRRCF-----VKVARYRSERL 765
Db 1685 FDEISDADKQFINL-LKKMKMRKLDCTQ---CLQHPWLMKOTKKNMKAELKSLDRMKKYM 1740
Qy 766 PSEOIGQVAVISVNLPRGRGFLSNPRAMGRFDSVTIGPNCACVPARCDOSPDAYSAYVL 825
Db 1741 ARKKWQ-----KTG--NAVRAIGRLSSMA-----MI 1764
Qy 826 ASLAGEELAVESSPEKFNPAIGVPOYLKLNRYRTHDEDDPVKKTARQISMARRPS 885
Db 1765 SGLSGRK-----STGSP---TSPLAEKLESEED--VSOAFLFAVAEKEKPHV 1807
Qy 886 AEESSNGPIYAFENLRACEAPPSAAHFRFYQIEGDYR-DYNTVPFNNED 932
Db 1808 KP-----YESKITRDLVEYEGSAARPD-CKTIG--YPRDEVVWFRRD 1846
```

RESULT 15  
TSP2\_HUMAN STANDARD; PRT; 1172 AA.  
AC P35442;  
DT 01-JUN-1994 (Rel. 29, Created)  
DT 01-JUN-1994 (Rel. 29, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE THROMBOSPONDIN 2 PRECURSOR.  
GN THBS2 OR TSP2.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RS SEQUENCE FROM N.A.  
RX MEDLINE=94010892; PubMed=8406456;  
RA Label T.L., Byers P.H.;  
RT "Sequence and characterization of the complete human thrombospondin 2



RT cDNA: potential regulatory role for the 3' untranslated region.";  
RL Genomics 17:225-229(1993).  
[2]  
RP SEQUENCE OF 560-1172 FROM N.A.  
RC TISSUE-Fibroblast;  
RX MEDLINE=92217961; PubMed=1559694;  
RA Label T.L., McCookey Mlewiec D.J., Distche C.M., Byers P.H.;  
"Thrombospondin II: partial cDNA sequence, chromosome location, and  
RT expression of a second member of the thrombospondin gene family in  
RL humans.";  
Genomics 12:421-429(1992).  
CC -1- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND  
CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN,  
CC LAMININ AND TYPE V COLLAGEN.  
CC -1- SUBUNIT: HOMOTRIMER, CROSS-LINKED BY DISULFIDE BONDS.  
CC -1- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.  
CC -1- SIMILARITY: CONTAINS 1 WFC DOMAIN.  
CC -1- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.  
CC -1- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.  
CC -1- SIMILARITY: CONTAINS 7 TSP TYPE-3 DOMAINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC at the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; L12350; AAA03703.1; -;  
DR EMBL; M81339; -; NOT\_ANNOTATED\_CDS.  
DR PIR; A42173; A42173.  
DR HSP; P00740; 11XA.  
DR MIM; 188061; -;  
DR InterPro; IPR000561; -;  
DR InterPro; IPR000884; -;  
DR InterPro; IPR001007; -;  
DR Pfam; PF00008; EGF\_2;  
DR Pfam; PF00090; TSP\_1; 3.  
DR Pfam; PF00093; WFC\_1;  
DR PROSITE; PS00022; EGF\_1; FALSE\_NEG.  
DR PROSITE; PS01186; EGF\_2; 1.  
DR PROSITE; PS00092; TSP1; 3.  
DR PROSITE; PS01208; WFC; 1.  
KW Glycoprotein; Cell adhesion; Calcium-binding; Heparin-binding; Repeat;  
KW EGF-like domain; Signal.  
FT SIGNAL 1 18  
FT CHAIN 19 1172  
FT DOMAIN 19 232  
FT DOMAIN 318 375  
FT DOMAIN 381 432  
FT DOMAIN 437 493  
FT DOMAIN 494 548  
FT DOMAIN 549 589  
FT DOMAIN 590 647  
FT DOMAIN 648 692  
FT DOMAIN 725 760  
FT DOMAIN 761 783  
FT DOMAIN 784 819  
FT DOMAIN 820 842  
FT DOMAIN 843 880  
FT DOMAIN 881 916  
FT DOMAIN 917 952  
FT DOMAIN 953 1172  
FT SITE 928 930  
FT DISULFID 266 266  
FT DISULFID 270 270  
FT DISULFID 553 564  
FT DISULFID 558 574  
FT DISULFID 577 588  
FT DISULFID 594 610  
FT DISULFID 601 619  
FT DISULFID 622 646  
POTENTIAL.  
THROMBOSPONDIN 2.  
HEPARIN-BINDING (POTENTIAL).  
WFC.  
TSP TYPE-1 1.  
TSP TYPE-1 2.  
TSP TYPE-1 3.  
EGF-LIKE 1.  
EGF-LIKE 2.  
CALCIUM-BINDING (POTENTIAL).  
EGF-LIKE 3.  
TSP TYPE-3 1.  
TSP TYPE-3 2.  
TSP TYPE-3 3.  
TSP TYPE-3 4.  
TSP TYPE-3 5.  
TSP TYPE-3 6.  
TSP TYPE-3 7.  
C-TERMINAL.  
CELL ATTACHMENT SITE (POTENTIAL).  
INTERCHAIN (PROBABLE).  
BY SIMILARITY.  
BY SIMILARITY.  
BY SIMILARITY.  
BY SIMILARITY.  
BY SIMILARITY.  
BY SIMILARITY.

FT DISULFID 652 665 BY SIMILARITY.  
FT DISULFID 659 678 BY SIMILARITY.  
FT DISULFID 680 691 BY SIMILARITY.  
FT CARBOHYD 151 151 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 316 316 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 330 330 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 457 457 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 584 584 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 710 710 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1069 1069 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 1172 AA; 129955 MW; 2AC7B8230E44C6F5 CRC64;  
  
Query Match 2.38; Score 143; DB 1; Length 1172;  
Best Local Similarity 20.88; Pred. No. 0.051;  
Matches 105; Conservative 56; Mismatches 189; Indels 154; Gaps 29;  
  
DB 86 CARPLRLART-----TDWTPAGSTQVYVHSPREGFWCLNREORPGQ 128  
DB 360 CASPFVFGECPCPSGLHSVDGEBSMPAEMTQCSVT-----CGSGTQONGR 406  
QY 129 NCSNYTVRFLECP-----DGLRDRTERI-----NSPSPMSKSCAAGCGOTGVOTRTRICL 178  
DB 407 SCDVTSNLTCLGPSIQTRACSLSKCDTRIRQDGMGSHGMSWSSCVTCG-VGNITRIRLCN 465  
QY 179 AEMVSL-----GSEASEEQHCKMGQDC-----TACDLTQPMG--QYNADCD--- 217  
DB 466 SPVPQMGSKNCKGSGREKACOGAPCPIDGRMSPWSPMSACTVTCAGGIRETRVCSPE 525  
QY 218 -----AC-----MC--ODFMLGAVSLP--GGAPASGAAYILTKTKPLLTOTDSD 259  
DB 526 PQYGGKACVGVQERQMKNSKCPYDGLSNPCFGACS-----SFPD 569  
QY 260 GREIRPGLCP-----DGKSLIKTKVKFAPVLYLMPKTSLSKAATIKAEFVAERP--YVW 312  
DB 570 GSWSC-GFCPVGFLGNGTCEBDLDECALVPDLCFSTSKVPCVNTQPGFHCIPCPRYRG 628  
QY 313 MNPETKARRAAGSVSLCCKATGKPRPDKYFWYHNTLDLPSTYKRESKLVLRKLDQHOAG 372  
DB 629 NQPVGVGLEAKTEKQVCEPE-NPCKDK--THN-----CHKHAECIYLGHFSDPM-- 675  
QY 373 EYFCKAQSDAGAVKSKVQLIYVTSDEFPCNPVPESYLLRPLPHDFONATNSFYDV--GR 431  
DB 676 -YKCECQTG-----YAGDGLICG--EDSDLDGPNLNLVCAITATFYICIDN 719  
QY 432 CPVKTCAGQODNGIIRCDAVONCC-----GISKTEREIQCSGYTLPTVAKKCS--- 481  
DB 720 CPHLPNSQGEPRD--KKGIDACDDDDNDNGV--TDKRD-NCQLLFNPRQADYKDEVG 773  
QY 482 --CORCTETRSIVKGRVSAADNGE 503  
DB 774 DRCDNCPYVHN--PAQIDTNNGE 795

Search completed: October 9, 2001, 11:53:54  
Job time: 219 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 9, 2001, 11:38:59 ; Search time 26.29 Seconds  
(without alignments)  
2730.270 Million cell updates/sec

Title: US-09-609-383-2

Perfect score: 6317  
Sequence: 1 MVGKAWVESFLVLEVTSL.....QSGVVASLRFPRVAQOPLIN 1184

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

#### Database :

A\_Geneseq\_0601:\*

- 1: /SIDS8/gcgdata/geneseq/geneseq/AA1980.DAT:\*
- 2: /SIDS8/gcgdata/geneseq/geneseq/AA1981.DAT:\*
- 3: /SIDS8/gcgdata/geneseq/geneseq/AA1982.DAT:\*
- 4: /SIDS8/gcgdata/geneseq/geneseq/AA1983.DAT:\*
- 5: /SIDS8/gcgdata/geneseq/geneseq/AA1984.DAT:\*
- 6: /SIDS8/gcgdata/geneseq/geneseq/AA1985.DAT:\*
- 7: /SIDS8/gcgdata/geneseq/geneseq/AA1986.DAT:\*
- 8: /SIDS8/gcgdata/geneseq/geneseq/AA1987.DAT:\*
- 9: /SIDS8/gcgdata/geneseq/geneseq/AA1988.DAT:\*
- 10: /SIDS8/gcgdata/geneseq/geneseq/AA1989.DAT:\*
- 11: /SIDS8/gcgdata/geneseq/geneseq/AA1990.DAT:\*
- 12: /SIDS8/gcgdata/geneseq/geneseq/AA1991.DAT:\*
- 13: /SIDS8/gcgdata/geneseq/geneseq/AA1992.DAT:\*
- 14: /SIDS8/gcgdata/geneseq/geneseq/AA1993.DAT:\*
- 15: /SIDS8/gcgdata/geneseq/geneseq/AA1994.DAT:\*
- 16: /SIDS8/gcgdata/geneseq/geneseq/AA1995.DAT:\*
- 17: /SIDS8/gcgdata/geneseq/geneseq/AA1996.DAT:\*
- 18: /SIDS8/gcgdata/geneseq/geneseq/AA1997.DAT:\*
- 19: /SIDS8/gcgdata/geneseq/geneseq/AA1998.DAT:\*
- 20: /SIDS8/gcgdata/geneseq/geneseq/AA1999.DAT:\*
- 21: /SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT:\*
- 22: /SIDS8/gcgdata/geneseq/geneseq/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6307	99.8	1184	21	AAV6657
2	6307	99.8	1184	22	Membrane-bound pro
3	6297	99.7	1184	20	Membrane-bound pro
4	3088.5	48.9	1156	20	AAW74445 (UNG)
5	2898.5	45.9	599	21	AAV23754
6	170	2.7	890	20	AAV49656
7	170	2.7	890	22	AAV49502
8	167.5	2.6	481	20	AAV50003
9	167.5	2.6	1072	22	AAV40145
10	160	2.5	905	22	AAV72300
11	157.5	2.5	1081	22	AAV72284
					Human ADAMTS-8 am
					Human ADAMTS-10 am

12	157	2.5	1152	21	AAV6657	Human thrombospond
13	157	2.5	1170	22	AAV6657	Human variant thro
14	156	2.5	680	21	AAV6657	Human metalloprote
15	155	2.5	1496	20	AAV6657	Melanoma associate
16	155	2.5	1496	21	AAV6657	Human p53 target m
17	154.5	2.4	4412	21	AAV6657	Sequence g1/101742
18	151.5	2.4	564	21	AAV6657	Human ACAM4/IgG4-F
19	151.5	2.4	598	21	AAV6657	Human ACAM4/IgG4-F
20	151.5	2.4	4393	22	AAV6657	Amino acid sequenc
21	148.5	2.4	300	21	AAV6657	Human COMP/TSP-2 c
22	147.5	2.3	1125	19	AAV6657	Rattus norvegicus
23	147.5	2.3	1139	19	AAV6657	Rattus norvegicus
24	144	2.3	1242	19	AAV6657	Rattus norvegicus
25	144	2.3	1911	16	AAV6657	Human ppp-OB
26	144	2.3	1911	18	AAV6657	Human ppp-OB
27	144	2.3	1911	20	AAV6657	Human protein tyro
28	143	2.3	1172	21	AAV6657	Human thymospond
29	143	2.3	1291	16	AAV6657	Tyrosine phosphata
30	142.5	2.3	4544	15	AAV6657	Alpha 2-Macroglobu
31	142.5	2.3	4544	15	AAV6657	Human alpha-2-MR
32	142	2.2	1172	21	AAV6657	Human thymospond
33	140.5	2.2	368	21	AAV6657	Human ACAM4/IgG1-F
34	140.5	2.2	398	19	AAV6657	A secreted protein
35	140.5	2.2	398	20	AAV6657	Human MBGP1 protei
36	140.5	2.2	398	20	AAV6657	Amino acid sequenc
37	140.5	2.2	398	21	AAV6657	Human ACAM cellula
38	140.5	2.2	398	21	AAV6657	Human ACAM cellula
39	140.5	2.2	398	21	AAV6657	Amino acid sequenc
40	140.5	2.2	398	22	AAV6657	Human LDCAM bindin
41	140.5	2.2	398	22	AAV6657	Human PRO258 prote
42	140.5	2.2	432	21	AAV6657	Human angiogenesis
43	140.5	2.2	432	21	AAV6657	Human ACAM cellula
44	140.5	2.2	432	21	AAV6657	Amino acid sequenc
45	140.5	2.2	433	21	AAV6657	Human LDCAM bindin
						Human Beat-1-like 1

#### ALIGNMENTS

RESULT 1	
AAV6657	standard; protein: 1184 AA.
ID	AAV6657
AC	AAV6657
XX	
DT	05-APR-2000 (first entry)
XX	
DE	Membrane-bound protein PRO1188.
XX	
KW	Membrane-bound polypeptide: PRO polypeptide; LDL receptor; TIE ligand;
KW	pharmaceutical; receptor immunoadhesin; gene mapping.
XX	
OS	Homo sapiens.
XX	
PN	W09963088-A2.
XX	
PD	09-DEC-1999.
XX	
PF	02-JUN-1999; 99WO-US12252.
XX	
PR	02-JUN-1998; 98US-0087607.
PR	02-JUN-1998; 98US-0087609.
PR	02-JUN-1998; 98US-0087759.
PR	03-JUN-1998; 98US-0087827.
PR	04-JUN-1998; 98US-0088021.
PR	04-JUN-1998; 98US-0088025.
PR	04-JUN-1998; 98US-0088028.
PR	04-JUN-1998; 98US-0088029.
PR	04-JUN-1998; 98US-0088030.
PR	04-JUN-1998; 98US-0088033.
PR	04-JUN-1998; 98US-0088326.
PR	05-JUN-1998; 98US-0088167.
PR	05-JUN-1998; 98US-0088202.

```
PR 05-JUN-1998; 98US-0088212.
PR 05-JUN-1998; 98US-0088217.
PR 09-JUN-1998; 98US-0088655.
PR 10-JUN-1998; 98US-0088722.
PR 10-JUN-1998; 98US-0088730.
PR 10-JUN-1998; 98US-0088734.
PR 10-JUN-1998; 98US-0088738.
PR 10-JUN-1998; 98US-0088740.
PR 10-JUN-1998; 98US-0088741.
PR 10-JUN-1998; 98US-0088742.
PR 10-JUN-1998; 98US-0088810.
PR 10-JUN-1998; 98US-0088811.
PR 10-JUN-1998; 98US-0088824.
PR 10-JUN-1998; 98US-0088825.
PR 10-JUN-1998; 98US-0088826.
PR 11-JUN-1998; 98US-0088858.
PR 11-JUN-1998; 98US-0088861.
PR 11-JUN-1998; 98US-0088863.
PR 11-JUN-1998; 98US-0088863.
PR 11-JUN-1998; 98US-0088876.
PR 12-JUN-1998; 98US-0089090.
PR 12-JUN-1998; 98US-0089105.
PR 16-JUN-1998; 98US-0089440.
PR 16-JUN-1998; 98US-0089512.
PR 16-JUN-1998; 98US-0089514.
PR 17-JUN-1998; 98US-0089532.
PR 17-JUN-1998; 98US-0089538.
PR 17-JUN-1998; 98US-0089538.
PR 17-JUN-1998; 98US-0089599.
PR 17-JUN-1998; 98US-0089600.
PR 17-JUN-1998; 98US-0089653.
PR 18-JUN-1998; 98US-0089801.
PR 18-JUN-1998; 98US-0089907.
PR 18-JUN-1998; 98US-0089908.
PR 19-JUN-1998; 98US-0089947.
PR 19-JUN-1998; 98US-0089948.
PR 19-JUN-1998; 98US-0089952.
PR 22-JUN-1998; 98US-0090246.
PR 22-JUN-1998; 98US-0090252.
PR 22-JUN-1998; 98US-0090254.
PR 23-JUN-1998; 98US-0090349.
PR 23-JUN-1998; 98US-0090355.
PR 24-JUN-1998; 98US-0090429.
PR 24-JUN-1998; 98US-0090431.
PR 24-JUN-1998; 98US-0090435.
PR 24-JUN-1998; 98US-0090444.
PR 24-JUN-1998; 98US-0090445.
PR 24-JUN-1998; 98US-0090461.
PR 24-JUN-1998; 98US-0090472.
PR 24-JUN-1998; 98US-0090535.
PR 24-JUN-1998; 98US-0090538.
PR 24-JUN-1998; 98US-0090540.
PR 24-JUN-1998; 98US-0090557.
PR 25-JUN-1998; 98US-0090676.
PR 25-JUN-1998; 98US-0090678.
PR 25-JUN-1998; 98US-0090690.
PR 25-JUN-1998; 98US-0090691.
PR 25-JUN-1998; 98US-0090694.
PR 25-JUN-1998; 98US-0090695.
PR 25-JUN-1998; 98US-0090862.
PR 26-JUN-1998; 98US-0090863.
PR 01-JUL-1998; 98US-0091358.
PR 01-JUL-1998; 98US-0091360.
PR 02-JUL-1998; 98US-0091544.
PR 02-JUL-1998; 98US-0091478.
PR 02-JUL-1998; 98US-0091486.
PR 02-JUL-1998; 98US-0091519.
PR 02-JUL-1998; 98US-0091626.
PR 02-JUL-1998; 98US-0091628.
PR 02-JUL-1998; 98US-0091633.
PR 02-JUL-1998; 98US-0091646.
PR 02-JUL-1998; 98US-0091673.

PR 07-JUL-1998; 98US-0091978.
PR 07-JUL-1998; 98US-0091982.
PR 09-JUL-1998; 98US-0092182.
PR 10-JUL-1998; 98US-0092472.
PR 20-JUL-1998; 98US-0093339.
PR 30-JUL-1998; 98US-0094651.
PR 04-AUG-1998; 98US-0095285.
PR 04-AUG-1998; 98US-0095285.
PR 04-AUG-1998; 98US-0095301.
PR 04-AUG-1998; 98US-0095302.
PR 04-AUG-1998; 98US-0095318.
PR 04-AUG-1998; 98US-0095321.
PR 04-AUG-1998; 98US-0095325.
PR 10-AUG-1998; 98US-0095916.
PR 10-AUG-1998; 98US-0095929.
PR 11-AUG-1998; 98US-0096102.
PR 11-AUG-1998; 98US-0096143.
PR 11-AUG-1998; 98US-0096146.
PR 12-AUG-1998; 98US-0096329.
PR 12-AUG-1998; 98US-0096757.
PR 17-AUG-1998; 98US-0096766.
PR 17-AUG-1998; 98US-0096768.
PR 17-AUG-1998; 98US-0096773.
PR 17-AUG-1998; 98US-0096791.
PR 17-AUG-1998; 98US-0096867.
PR 17-AUG-1998; 98US-0096891.
PR 17-AUG-1998; 98US-0096894.
PR 17-AUG-1998; 98US-0096895.
PR 17-AUG-1998; 98US-0096897.
PR 18-AUG-1998; 98US-0096849.
PR 18-AUG-1998; 98US-0096950.
PR 18-AUG-1998; 98US-0096959.
PR 18-AUG-1998; 98US-0096960.
PR 18-AUG-1998; 98US-0097022.
PR 19-AUG-1998; 98US-0097141.
PR 20-AUG-1998; 98US-0097218.
PR 24-AUG-1998; 98US-0097661.
PR 26-AUG-1998; 98US-0097951.
PR 26-AUG-1998; 98US-0097952.
PR 26-AUG-1998; 98US-0097954.
PR 26-AUG-1998; 98US-0097955.
PR 26-AUG-1998; 98US-0097971.
PR 26-AUG-1998; 98US-0097974.
PR 26-AUG-1998; 98US-0097978.
PR 26-AUG-1998; 98US-0097979.
PR 26-AUG-1998; 98US-0097986.
PR 26-AUG-1998; 98US-0098014.
PR 31-AUG-1998; 98US-0098525.
PR 16-SEP-1998; 98US-0100634.
PR 12-JAN-1999; 99US-0115565.

PA (GETH ) GENENTECH INC.
XX
XX
PI Baker K, Chen J, Goddard A, Gurney AL, Smith V, Watanabe CK;
PI Wood WI, Yuan J;
XX
XX
XX MPI: 2000-072883/06.
XX N-PSDB; AAZ64985.
XX
XX Membrane-bound proteins and related nucleotide sequences
XX
XX claim 12; Fig 72; 822pp; English.
XX
XX The invention provides membrane-bound PRO polypeptides and
XX polynucleotides encoding them. The PRO sequences of the invention were
XX identified based on extracellular domain homology screening. The PRO
XX sequences have homology with proteins including LDL receptors, TIE
XX ligands and various enzymes. The membrane-bound proteins and receptor
XX molecules are useful as pharmaceutical and diagnostic agents. Receptor
XX immunoadhesins, for instance, can be used as therapeutic agents to block
XX receptor-ligand interactions. The membrane-bound proteins can also be
XX employed for screening of potential peptide or small molecule inhibitors
XX of the relevant receptor/ligand interaction. The PRO encoding sequences
```

are useful as hybridization probes, in chromosome and gene mapping and in the generation of antisense RNA and DNA. PRO nucleic acid sequences will also be useful for the preparation of PRO polypeptides, especially by recombinant techniques.

Sequence 1184 AA;

Query Match	99.8%;	Score 6307;	DB 21;	Length 1184;
Best Local Similarity	99.8%;	Prod. No. 0;	Mismatches 1182;	Conservative 0;
Matches 1182;	Conservative 0;	Mismatches 2;	Indels 0;	Gaps 0;
OY	1	MVGTAWFSEFVLEVTSLGKQMTQSVRRVOPGKKNPSTFAKPADLSEPGEMTTLF	60	
DB	1	mvgtawfsefvlvlevtslgkqmtqsvrrvopgkknpsfakpadllespgemtlf	60	
OY	61	NIDYBGKGDYERLDAIRFYVDRCARPLRLAARTDWTAGSTGOVHSGPRGFWCL	120	
DB	61	nidybgkgdyerldaifryvdrcarplrlaartdwtltpgstgqvvhsgprgfwcl	120	
OY	121	NREORPQNCNVTYRFLCPGSLRDRTERLWSPWSPWKSACACGQTEGOTRTRICAE	180	
DB	121	nreorpgncsnvtvrlcpgslrdrterlwspwspwksacaacqtegvtrtriciae	180	
OY	181	MVSLCSASESGOHMGDDCTACDLTCPMGOVNADCDACMCDPMLHGAVSLPGAPASG	240	
DB	181	mvslcsaseeeghmgddctacdltpcmgynadcdacmqdflmgavslpgapasg	240	
OY	241	AAIYLTTPKLLTQTDSDGRFRIPGLCPDGKSLKTKTVKAPAPVLMMPKSLKAATIK	300	
DB	241	aaiyltltkplltqtdsdgrfrlpglcpdgksllkltkvkfpaplvlmpktslkaatik	300	
OY	301	AEFVRAETPYWYMNDETKARRAGGSVSLCCAKATGKPRDQKYFWYHNDLLDPSLYKHESK	360	
DB	301	aefvraetpywmnpetkarraggsvsllccakatkprdkyfwymndllldpslykhesk	360	
OY	361	LVLRKLOOHGGEYFCCKKOSDGAVKRYAOLIVTASDTPCNPPVESTLRLPHDCPON	420	
DB	361	lvlrkloohgeyfcckkksosdagavksyaolivtasdtpcnppvesylrlphdcpon	420	
OY	421	ATNSFYVVGRCPVTCAGQODNGIRCDRAVONCCGISKTEREIQSGSYTLPTFYAKEC	480	
DB	421	atnsfyvdvgrcprvtcagqodnglrcrdavngccgiskteereiqcsytlptkyakec	480	
OY	481	SCQCTETRSIYVRGVSAAADNGEPMRFGHVWGSRVSMGTGKGPFTLHVPODTERLYLT	540	
DB	481	scqctetrsiyrvgvsaaadngemrfghvwnsrmvsmgtgkgtftlhvpofterlylt	540	
OY	541	FVDRLOKEVNTTKVLPFKKKGSVAFHEITKMLRRKEPITLEAMETNITPLGEVVGEDPMAE	600	
DB	541	fvdrloketvnttkvlpfkkgksavfheitkmlrrkepitleamethlplgevvgedpmae	600	
OY	601	LEIPRSFYRONGEPIYIGKVASVTFELDPRI NSTATAOTDNFNINDEGDTPLRTYGMF	660	
DB	601	leiprsfyrgngepylqkvavsvtfelprlnstataotdnfnindegdtplrtlygmf	660	
OY	661	SVDFDEVTSSEPLNAGKVKVHLDSQVKMPEHISTVKLWSLNDPGLWEEBGEDEFENOR	720	
DB	661	svdfdevtseplnagkvyvhldsqvkmpehistsvklwslndpdlweeegdefenfgr	720	
OY	721	RKKRRDRFTVLGNLRIERRLFNLDVPESRRCFVVRATRSRRLPSEDIQGVISVTL	780	
DB	721	rkkrrdrftvlgnlrierrlfnldvpesrrcfvvratrslrslpseiqgvvisvltl	780	
OY	781	EBPRTGFLSNPRAMGRPDSVITGPNACVAFCDOSPDVASYAVLASLAGELOAVESSP	840	
DB	781	eprtgflsnpramgrpdsvitgpnacvafcdospdvasyavlaslagedlavessp	840	
OY	841	KKNPNAIGVPOPYLKNLNRTRTDHEDPRVKTAFQISMAKPPNSAESNGPIYAFENLR	900	
DB	841	kknpnaiigvppylknlnrtrtdhedprvktafqismakppnsaesngpiyafenlr	900	
OY	901	ACEEAPPSAAHRRFYQIBEDRDYDNTVPPNEDDPMKMTEDYLAWMPKPMERFACITIKYKI	960	

DB	901	aceeapshaahrrfyqibedrdydyntvppneddpmkmtedylawmpkpmefraciyikvki	960	
OY	961	VGPLEVNVRSRNMGGTHRRRTVGLKGIKIDVNRSTRDPOPNVSAACLEFKCSMLYDDORV	1020	
DB	961	vgplevnvrsmggthrrrtvglkgiidvnrstrdpoPNVsaaclefkcsgmlyddqdrv	1020	
OY	1021	DRTLVKVIPOGSCRRASVPMIHEYLNVNRLPLAVNNDSSEYMLAFLDPLGHNVGILYVT	1080	
DB	1021	drtlvkvipogscrrasvpmihelylvnrlplavnndsseymtlaflpdlghnygilyvt	1080	
OY	1081	DODPRTAKEIALGRCFDGTSDGSSRIMKSNMGVALTFNCVERQVGRQSAFOYLOSTPAOS	1140	
DB	1081	ddprrtakeialgrcfdgtdsgssrimsnmgvaltfncverqygrqsafgyldgstpas	1140	
OY	1141	PAAGTVQGRVPSRROORASRGOSGVVASLRPPRVAOOPLIN 1184		
DB	1141	paagtvqgrvpsrrqrasrgqrgyvaslrfprvaqplln 1184		
RESULT 2				
ID	AAB65180	standard; Protein; 1184 AA.		
XX				
AC	AAB65180;			
DT	02-APR-2001 (first entry)			
XX				
DE	Human PRO1188 (UNQ602) protein sequence SEQ ID NO:124.			
XX				
KN	Human; secreted and transmembrane protein; PRO; cytosolic;			
KW	cell death; cancer; chromosomal mapping; gene mapping; tissue typing;			
XX	diagnostic assay.			
OS	Homo sapiens.			
XX				
PN	WO200073454-A1.			
PD	07-DEC-2000.			
XX				
PE	30-MAR-2000; 2000WO-US08439.			
PR	02-JUN-1999; 99WO-US12252.			
PR	23-JUN-1999; 99US-0141037.			
PR	07-JUL-1999; 99US-0143048.			
PR	20-JUL-1999; 99US-0144758.			
PR	26-JUL-1999; 99US-0145698.			
PR	28-JUL-1999; 99US-0146222.			
PR	17-AUG-1999; 99US-0149396.			
PR	15-SEP-1999; 99WO-US21090.			
PR	15-SEP-1999; 99WO-US21547.			
PR	08-OCT-1999; 99US-0158663.			
PR	30-NOV-1999; 99WO-US28313.			
PR	01-DEC-1999; 99WO-US28301.			
PR	16-DEC-1999; 99WO-US30095.			
PR	20-DEC-1999; 99WO-US30911.			
PR	05-JAN-2000; 2000WO-US00219.			
PR	06-JAN-2000; 2000WO-US00376.			
PR	11-FEB-2000; 2000WO-US03565.			
PR	18-FEB-2000; 2000WO-US04341.			
PR	22-FEB-2000; 2000WO-US04414.			
PR	24-FEB-2000; 2000WO-US04914.			
PR	24-FEB-2000; 2000WO-US05004.			
PR	15-MAR-2000; 2000WO-US05844.			
PR	20-MAR-2000; 2000WO-US07377.			
PA	(GETH ) GENENTECH INC.			
XX				
PI	Ashtkenazi AJ, Baker KP, Botstein D, Desnoyers L, Eaton DL;			
PI	Ferrara N, Fong S, Gerber H, Gerlitsen ME, Goddard A, Godowski PJ;			
PI	Grimaldi CJ, Gurney AU, Kljavin IJ, Napier MA, Pan J, Paoni NF;			
PI	Roy MA, Stewart TA, Tumas D, Watanabe CK, Williams PM, Wood WT;			

PI Zhang Z;  
 XX  
 DR WPI; 2001-032160/04.  
 DR N-PSDB; AAF44131.  
 XX  
 PT PRO polynucleotides used to produce polypeptides used to target  
 PT bioactive molecules such as toxins, radiolabels or antibodies, to  
 PT specific cells, to cause targeted cell death -  
 PS  
 PS Claim 12; Fig 72; 935pp; English.  
 XX  
 CC The present invention describes human secreted and transmembrane PRO  
 CC proteins. The PRO proteins have cytostatic activity. The PRO proteins  
 CC can be used for targeted delivery of bioactive molecules, such as  
 CC toxins, radiolabels or antibodies, that cause cell death. PRO nucleotide  
 CC sequences, and their fragments, can be used as hybridisation probes, in  
 CC chromosomal and gene mapping, and in the generation of anti-sense RNA  
 CC and DNA. They may also be used to produce transgenic animals which are  
 CC used to develop and screen therapeutically useful reagents. The PRO  
 CC nucleotide and protein sequence can be used for tissue typing and in  
 CC treating cancer. Anti-PRO antibodies can be used in diagnostic assays.  
 CC AAF44270 to AAF44470 represent PCR primers and hybridisation probes used  
 CC in the isolation of human PRO sequences. AAF4087 to AAF44269 and  
 CC AAB65154 to AAB65300 represent human PRO polynucleotide and protein  
 CC sequences given in the exemplification of the present invention.  
 CC  
 XX  
 SQ Sequence 1184 AA;  
 Query Match 99.8%; Score 6307; DB 22; Length 1184;  
 Best Local Similarity 99.8%; Pred. No. 0;  
 Matches 1182; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 MVTGKAWFSEFLVLEVTSVLGRQTMLTQSVRVOPGKKNPSIFAKPADTLESPEWTLF 60  
 1 mvgkavwfsflvleevsvlgrqtmlltqsvrvopgkknpsifakpadtlespewtlf 60  
 QY 61 NIDYPGKGDEYERLDATRFYIGDRVCARPLLEARTDWPAGSTGOVHGSPREGWCL 120  
 61 nidypgkgderyerldatrfyigdrvcarplleartdwpagstgovhgspregwcl 120  
 Db 61 nidypgkgderyerldatrfyigdrvcarplleartdwpagstgovhgspregwcl 120  
 QY 121 NREORPGONCNVTYVRFCLPSPGSLRDRTERIWSMSPMSKSAACGGOTFRIRICAE 180  
 121 nreorpgoncnvtyvrfclpspgslrdrteriwsmspsmsksaacggotfriricae 180  
 Db 121 nreorpgoncnvtyvrfclpspgslrdrteriwsmspsmsksaacggotfriricae 180  
 QY 181 MVSLSSEASESGOHOMQDCTACDITCPMGVNAQDCAMQODMLHGAASLPGAPASG 240  
 181 mvsclsseasegohomqdctacdltcpmgvnaqdcamqodmlhgaaslpgapasg 240  
 Db 181 mvsclsseasegohomqdctacdltcpmgvnaqdcamqodmlhgaaslpgapasg 240  
 QY 241 AATYLLKTRKRLTQTSDSGRFRIPGLCPDGSKILKTKKFAPIVLTMPKTSKATIK 300  
 241 aatyllktrkrltqtsdsgrfripglcpdgskilktkkfapiylvltmpktskaatik 300  
 Db 241 aatyllktrkrltqtsdsgrfripglcpdgskilktkkfapiylvltmpktskaatik 300  
 QY 301 AEFRAETPYWVWNPETKARAGOSVSLCCATGKPRDXFEWHNPILDPISLYKHESK 360  
 301 aefraetpywvwnpetkaragosvslccatgkprdxfewhnpildpisllykhesk 360  
 Db 301 aefraetpywvwnpetkaragosvslccatgkprdxfewhnpildpisllykhesk 360  
 QY 361 LVLRKLOHQAQGEYFCKAQSDAGAVKSKVAQLIVTASDEPCNVPESYLIRLPHDFON 420  
 361 lvlrklqhgaqgeyfckagsdagavkskvaqlivtasdepcnvpesylirlphdfon 420  
 Db 361 lvlrklqhgaqgeyfckagsdagavkskvaqlivtasdepcnvpesylirlphdfon 420  
 QY 421 ATNSFYDVGRCPRYKTCAGQDNGIRCHDAVQNCGISKTEREIQSGYTLPRKVAKEC 480  
 421 atnsfydvgrcpryktcagqdngrchdavnqncgisaktereiqsgytlprkvahec 480  
 Db 421 atnsfydvgrcpryktcagqdngrchdavnqncgisaktereiqsgytlprkvahec 480  
 QY 481 SCORCTEIRSVIRGRVSAADNGEPMRFGHYVNGNSRVSMTGKGFTHVPODERLVL 540  
 481 scortetirsvirgrvsaadngepmrfghyvmgnsvsmgkykgtfthvpoederlvt 540  
 Db 481 scortetirsvirgrvsaadngepmrfghyvmgnsvsmgkykgtfthvpoederlvt 540  
 QY 541 FVDRLOQFVNTTKVLPFNKKSAAVFEHMKLRREKPTTLEAMETNIIPLGEVGEDPMAE 600  
 541 fvdrlqfvtntkvlpfnkksaavfehmklrrekepittleametniiplgevgedpmae 600  
 Db 541 fvdrlqfvtntkvlpfnkksaavfehmklrrekepittleametniiplgevgedpmae 600

QY 601 LEIPSRSEFYRQNGCEPIYIGKVASVTFLEDPNRNISTATAQTDLINEDEGDTPLRTYGMF 660  
 601 leiprsrsefyrqngcepiyigkvasvtfledpnrnistataqtdlinedegdtplrtymf 660  
 Db 601 leiprsrsefyrqngcepiyigkvasvtfledpnrnistataqtdlinedegdtplrtymf 660  
 QY 661 SVDFRDEYTSSEPLNAGKVKVHLSDSTQVKMPHEHISTVKLMSINPDTGIMEEGDFKFNOR 720  
 661 svdfredeytsseplnagkvkvhlstdstqvkmphehistvklmsinpdtdgimeegdfkfnor 720  
 Db 661 svdfredeytsseplnagkvkvhlstdstqvkmphehistvklmsinpdtdgimeegdfkfnor 720  
 QY 721 RNKREDRTFLVGNLEIEERRLEFNLDVPESRCFYKVAASREERFLPSEOIGQVTSVINL 780  
 721 rnkredrtflvgnleieerrlefnldvpesrcfykvaasreerflpseoirgqvtsvinl 780  
 Db 721 rnkredrtflvgnleieerrlefnldvpesrcfykvaasreerflpseoirgqvtsvinl 780  
 QY 781 EPRTGFLSNPRAMGRFDSVTITGPNAGACVPAFCDDQSDPAYSAVYLASLAGELOAVESSP 840  
 781 eprtgflsnpramgrfdsvtitgpnagacvpafcdqsdpaysavylaslageleavessp 840  
 Db 781 eprtgflsnpramgrfdsvtitgpnagacvpafcdqsdpaylavylaslageleavessp 840  
 QY 841 KFNNAIGVPPYLNKLNKRRTHDEDRVKKTAQISMAKRRPSAEESSNCPITAFENLR 900  
 841 kfnnaigvppylnklnkrrthdedrvkktaqisamakrrpsaeesncpitaefenlr 900  
 Db 841 kfnnaigvppylnklnkrrthdedrvkktaqisamakrrpsaeesncpitaefenlr 900  
 QY 901 ACEBAPPSAAHFRFYQIEGRDYDNTVPENEDPMSTEDYLAWMPKPMERACIYKVI 960  
 901 acebappsaahfrfyqiegrdydntvpenedpmstedyllawmpkpmetracyikvi 960  
 Db 901 acebappsaahfrfyqiegrdydntvpenedpmstedyllawmpkpmetracyikvi 960  
 QY 961 VGPLEVNVRSNMGTGHRRTVGLYGRDVSTRDQPNVSAACLEFKSGMLYDODRV 1020  
 961 vgplevvnvrsnmgtghrrtvglgrdvstrdqpvnvsaaclefksgmlydodrv 1020  
 Db 961 vgplevvnvrsnmgtghrrtvglgrdvstrdqpvnvsaaclefksgmlydodrv 1020  
 QY 1021 DRTLKVTYPOGSCRRASVNPMLHEYLNVHLPLAVNPDSEYTMALPDPGHNGITVY 1080  
 1021 drtlkvtypogscrrasvnpmlheylnvhlplavnpdseytmalpdpgnhngitvy 1080  
 Db 1021 drtlkvtypogscrrasvnpmlheylnvhlplavnpdseytmalpdpgnhngitvy 1080  
 QY 1081 DQDPRTAKEILAGRCFQDTSDDSGSRIMKSNVGVALLFNCVERQYGVROSAQYLOSTPAOS 1140  
 1081 dqdprtakeilagrclfqdtssdgsrimsnvgvallfncverqygvrosaqylostpas 1140  
 Db 1081 dqdprtakeilagrclfqdtssdgsrimsnvgvallfncverqygvrsaqylostpas 1140  
 QY 1141 PAAGTVGGRVPSRQQRASRGQROSGVVASLRPPRYAQQPLIN 1184  
 1141 paagtvgrvpsrrqqrassrgqrosgvvaslrppryaqqpln 1184  
 Db 1141 paagtvgrvpsrrqqrassrgqrosgvvaslrppryaqqpln 1184

RESULT 3  
 AAM74445  
 ID AAM74445 standard; Protein: 1184 AA.  
 XX  
 AC AAM74445;  
 XX  
 DT 12-MAY-1999 (first entry)  
 XX  
 DE Human nucleotide pyrophosphohydrolase, NTPPH-1.  
 XX  
 KW NTPPH-1; human; nucleotide pyrophosphohydrolase; arthropathy; therapy;  
 KW immunological disorders; cancer; haemodialysis; infection;  
 KW extracellular circulation.  
 XX  
 OS Homo sapiens.  
 XX  
 PN US5876963-A.  
 XX  
 PD 02-MAR-1999.  
 XX  
 PF 27-AUG-1997; 97US-0918914.  
 XX  
 PR 27-AUG-1997; 97US-0918914.  
 XX  
 PA (HUTC/) HUTCHINSON N.  
 PA (LAWT/) LAWTON M.  
 PA (MAGN/) MAGNA H.  
 PA (MITC/) MITCHELL P.  
 PA (MURR/) MURRY L E.  
 PA (YOCU/) YOCUM S.  
 XX  
 PI Hutchinson N, Lawton M, Magna H, Mitchell P, Murry LE;

PI Yocum S;  
 XX WPI: 1999-189634/16.  
 DR N-PSDB; AAX18449.  
 XX  
 PT New human nucleotide pyrophosphohydrolase - useful for providing  
 PT methods for identifying and treating arthropies, immunological  
 PT disorders, and cancer  
 XX  
 PS Claim 1; Fig 1; 42pp; English.  
 XX  
 CC This sequence represents the human nucleotide pyrophosphohydrolase,  
 CC designated NPPH-1, of the invention. NPPH-1 antagonists, antibodies,  
 CC agonists, proteins, complementary sequences or vectors can be used to  
 CC treat and identify arthropies (e.g. calcium pyrophosphate dihydrate  
 CC deposition disease, degenerative joint disease, fibromyalgia,  
 CC haemochromatosis, osteoarthritis, progressive systemic sclerosis,  
 CC pseudogout, psoriasis, rheumatoid arthritis and lupus erythematosus);  
 CC immunological disorders (e.g. AIDS, allergies, anaemia, asthma,  
 CC ulcerative colitis, dermatomyositis, diabetes mellitus, emphysema,  
 CC glomerulonephritis, gout, multiple sclerosis, osteoporosis and  
 CC pancreatitis); trauma; complications of cancer, haemodialysis, and  
 CC extracorporeal circulation; viral, bacterial, fungal, parasitic,  
 CC protozoal, and helminthic infections; and cancer (e.g. adenocarcinoma,  
 CC lymphoma, melanoma, myeloma, sarcoma, leukaemia, or teratocarcinoma of  
 CC the bone and bone marrow, brain, breast, cervix, gastrointestinal tract,  
 CC kidney, liver, lung, ovary, testis and skin).  
 XX  
 SQ Sequence 1184 AA;  
 Query Match 99.7%; Score 6297; DB 20; Length 1184;  
 Best Local Similarity 99.7%; Pred. No. 0;  
 Matches 1181; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

DB 541 fvdrlgkfnttclvlpfnkkgasvfhelkmlcrkepltleametiilpigevedpmae 600  
 QY 601 LEIPSRSEYRQNGEPIYIGKVASVTFLEDPRIISTATAQTDLNFTINDGDTPLPTTGMF 660  
 DB 601 leiprsfyrnggepylgkvkavstfldprnistaetqtdlnfndegdtlptlygmf 660  
 QY 661 SVDFDEYTSSEPLNKGKVVHLDSTQVKKPEHISVVKLMSLNPDGIMWEEEDGKFEENOR 720  
 DB 661 svdfdevtseplnagkvkvhldstqvkkpehislvslnpdgllweeedgkfeenr 720  
 QY 721 RNKREDRTFLVGNLEIRERRFLNLDVPESRRCFVVRVARSERFLPSEQIGVSVIYL 780  
 DB 721 rnkredrtflvgnlerrflnldvpesrtrcfvvraserflpseqigvsviyl 780  
 QY 781 EPRTFPLSNPRAMGRDYSYITGPNACVPACDDOSPDAYSATVLAAGELAVESSP 840  
 DB 781 eprtlflsnpramgrdysyitgpnacvpafcdodspdaysavlaslaageelavessp 840  
 QY 841 KPNPAIGVPOPLYLKLNVNRRTHEDDPYKTAFOISMAKPPNPAESNGPIYAFENLR 900  
 DB 841 kpnpaigvppqpylnklnvrrthedpvyktafaismakppnpsaesngpiyafenlr 900  
 QY 901 ACEEAPPSAHRFRFYQIEGDRYNTVPFNEEDPMKMTEDYLAWMPKPMEEFRACYIKVKI 960  
 DB 901 aceeappsaahrfrfyqiegdryntvpfneedpmkmtedyawmpkpmefracyikvki 960  
 QY 961 VGPLEVNVRSRMGGTHRRTVGKLGIRDVNSTRRDQPNVSACLERKCSGMLYDDDRV 1020  
 DB 961 vgplevnrsmggthrrtvgklgirdvnstrrdqpnvsaclefkcsgmlyddqdrv 1020  
 QY 1021 DRTLVKVIPOGSCRASVNPMLHEYLNVNHLPLAVNNDSSEYFMLAPLOPLGNVGIYPT 1080  
 DB 1021 drtlvkvipogscrrasvnpmlheylvnhlplavnndsseymldplgnmygiylvt 1080  
 QY 1081 DDDPRFAKIALGRCFDGTSDGSSRIMKSNVVALTFMNCVEROVSQSAFOYLQSTPAQS 1140  
 DB 1081 dddprtakeialgrcfdsdgsrilmksnvvaltfmncvervgysafoylqstpqas 1140  
 QY 1141 PAAGTVQGVNPSRRROORASRGOROSGVYASLRFPRVAQOPLIN 1184  
 DB 1141 paagtvqgvnpsrrqrasrgqrsqyvaslrfprvaqopliln 1184

RESULT 4  
 AAY23754  
 ID AAY23754 standard; Protein; 1156 AA.  
 XX  
 AC AAY23754;  
 XX  
 DT 09-SEP-1999 (first entry)  
 XX  
 DE Human nucleotide pyrophosphohydrolase-2.  
 XX  
 KW Human; nucleotide pyrophosphohydrolase-2; NPPH-2; arthropathy;  
 KW Bechet's syndrome; degenerative joint disease; lupus erythematosus;  
 KW rheumatoid arthritis; psoriasis; immunological disease;  
 KW acquired immune deficiency syndrome; Addison's disease;  
 KW ulcerative colitis; osteoporosis; microbial infection; cancer.  
 XX  
 OS Homo sapiens.  
 OS  
 PN WO9932610-A1.  
 PN  
 PD 01-JUL-1999.  
 PD  
 PF 02-DEC-1998; 98WO-US25558.  
 PF  
 PR 22-DEC-1997; 97US-0996083.  
 PR  
 PA (INCY-) INCYTE PHARM INC.  
 PA  
 PI Hutchinson N, Lawton M, Magna H, Mitchell PG, Murry LE;  
 PI Schaffer P, Yocum SA;

XX WP1: 1999-4181919/35.  
DR N-PDSB: AAX81946.

Human nucleotide pyrophosphohydrolase-2

XX  
PS  
XX Claim 1; Fig 1A-K; 89pp; English.

The present sequence represents human nucleotide pyrophosphohydrolase-2 (NPPH-2). Antagonists of the protein are used to treat arthropathy (e.g. Behcet's syndrome, degenerative joint disease, lupus erythematosus, rheumatoid arthritis, psoriasis and many others); immunological disease (e.g. acquired immune deficiency syndrome, Addison's disease, ulcerative colitis, osteoporosis, microbial infections and many others), also a wide range of cancers.

XX  
XX Sequence 1156 AA;

Query Match 48.9%; Score 3088.5; DB: 20; Length 1156;  
Best Local Similarity 50.3%; Pred. No. 6.5e-271;  
Matches 584; Conservative 177; Mismatches 360; Indels 39; Gaps 12.

DQ 24 TMLTGSVRRVPGGKNKPSIFAKPADLTSPGEWTLFNIDYGCGKDRELDALRFYYCD 83  
Dd 32 talglerrsvtyqgspal-----edweasewtsfivndhyggsgdfeslaifffy9p 86  
QY 84 -RYCARPLRLERATDTWDTPAGSTGVVHGSPREGFWCLNRQRPQNCSNTYVFRELCPG 142  
Dd 87 arvcprrplaeattctdwalpssavgervhlnprgfwtclnregpyrcsnynhvrlrccp-- 144  
QY 143 SLKRRTETIWSWMSWMSKCSAACGTGYQTRIRICLAEKVSLICSEASLEGCHMGQDCTA 202  
Dd 145 -----leaswagawgwpgwcscgc-pqrtrlrirrhpspadacparpleaqkcvrrpcp 198  
QY 203 CDLTCPMGQVANADOCACMCODFMLHGAVSLPBGAPASGAIAITLKTKPKLLQTQSDGRF 262  
Dd 199 csl-----dtcecpdhlllgsvvpcpgqlllgavslrdq-pgvatsahngft 246  
QY 263 RIPGLCPDGKSILKITKYKF-APIVLTPWPKUSLKAAITKAEFRAETPMVANNPETKAR 321  
Dd 247 rvpvcadstranxrxmqmfgsaegdaqangisavltilld-klekplyvhpesryve 304  
QY 322 AGQSLSLCCKATGKRPRDKTYFYANDTLIDSPLYKHESKLVRKUQHOGAGEFYCKAOSD 381  
Dd 305 aggnvtfcckasqcpmpkykswfihngtllidrtabhygahelelrglrpdpaqiyhkawe 364  
QY 382 AGAVSKYAOLIVTASDEPCPNVPESYLIRPHHCFFQNATSIFYVDYGRGCYVKTCAGQG 441  
Dd 365 agavasygtarlvtlapgsbacoprpelylkripedcggqgsppaylldvglpdcitcpsla 424  
QY 442 DNGIRCDAVONCCGISKTEREIOCSGYTLPTKAKKECSCORCYETETSIVAGRYSADN 501  
Dd 425 gspiccgassrcscvrlerrelncpvytlpvkvaaecgcqklcprjgylvgrvaads 484  
QY 502 GEPKMFHVYWGNSVSMATGKTFTLTVHPDDTELVLFVDRLOKFEVNTTVLPFPNNKG 561  
Dd 485 gepilfarilligpedigtlaygdftilevppstqlvtlfdvbpsgefmadvavllpdfpg 544  
QY 562 SAVPEIKMLRKBEITLLEAMETNIIPIGEVYGEPEMELRIPSFSYFRQNEEPIYGIWK 621  
Dd 545 agyvnevkaamrkcapvllhtsgsncliprigeledapbelvipsqaftradqkxpsyape 604  
QY 622 ASVFLFDLRNRISTATAAQDTDLNFINDCEDTFLRIYGFSGFVSVDPRDVESEPINACKVKVH 681  
Dd 605 avrtltvdprdltaasapsodrlfvsdgelaprltrygmfsvvlirpssaeqqlyqvavyr 664  
QY 662 LDSTOVKKMEHIISTYKLMSLANDPTGLMBEGDFKENOR--RNKREDRTFLVGNLEIRER 739  
Dd 665 vaasgilmpghvwaalkwlslnpetjlweesgpfirregsgpvrvtreeryflvgaveiter 724  
QY 740 RLFNLDVDESRCPFKVAAYRSERLPLSFOLOGVIISYNLEPRKGLSNPAKMRPDSY 799  
||||||| ||||||| :||| |||||:::||| | | :||| ||| |||

```
Dd      725 rlinldperricrcvkrtayandkfipbsqevgvvvltlnlnebpagfsanprtaxwrfdsaa 784
Oy      800 ITGPNACVPAFCDDQSPDAYSAVVYLAISLAGEELQAVESSPKFNPNAIGVPDPYLNTKLNYY 859
Db      785 vtpgnacplfcdadrpdaytalvratlvatleapapslpplrlptvgvtppjdydlrlyg 844
Oy      860 RRTIHEDPRVKATIAFOISMAMPNPSAEEBSNGPIYAFENLRACEAPSAHFREFYQIEG 919
Dd      845 rtidhdpafrknrgftinlakprpgdaeanangpywrslrecqagpvashitfarvea 904
Oy      920 DRYOTNYVFNEHDEDDPMWTEJDLYAWMPKPMEHFACITAKTIVGLPELVANRKNMGHTIR 979
Dd      905 dkcyeynvvpifregpsawtgdllawmpnpefractlikvkikqbgelywmshagnshpr 964
Oy      980 TVGKITGYRDVRSTRHDMDQNVNSAACLEFFCSGMTKDODRVDRTLTKVIIPQSCRRASVN 10339
Dd      965 trtgqlgtlarasvrdepbergsataacevfacsqmlfdqrvqdrlvltmpggscrrcvaav 1024
Oy      1040 PMLHEIYVNHPLPLAVNNDDSEMYTMLPDLDTLGNNCYTYVTODDPTFAKEINLAGCPGT 1099
Dd      1025 gltltdyltrhpppyaeedpaafsmilaplddlghnyvyvrltdqsplakelaigtcfids 1084
Oy      1100 SDGSSTRIKSNVGVALTFNFCVEROVGRQSOFYOJLOSTPASOPAGTVGOGRVPSRRQOAS 1159
Dd      1085 sdcgfstrmkkaagtavrflgcrcprepagrpslfqrlllespa--taigdir-----remseaa 1137
Oy      1160 RGGOROSGVNASLFEPVAQQ 1179
Dd      :   | : | : : | |||
        db 1138 qaqaragaajplrtr-grvrg 1156
```

XX	RESULT	5
XX	AA49966	
ID	AA49966	standard; Protein; 599 AA.
XX		
AC	AA49966;	
XX		
DT	03-FEB-2000	(first entry)
XX		
DE	Porcine chondrocyte nucleotide pyrophosphohydrolase.	
XX		
KW	Porcine; pig; chondrocyte; nucleotide pyrophosphohydrolase; NTPPH;	
KW	hydrolysis; nucleoside triphosphate; ATP; nucleoside monophosphate;	
KW	ppi; calcium pyrophosphate dihydrate; CPPD; articular cartilage;	
KW	age; articular tissue degeneration; arthritis; joint tissue breakdown;	
KW	vesicle formation; diagnosis; cartilage.	
XX		
OS	Sus scrofa.	
XX		
PN	US5986080-A.	
XX		
PD	16-NOV-1999.	
XX		
PF	17-OCT-1997; 97US-0954333.	
XX		
PR	18-OCT-1996; 96US-0028702.	
XX		
PA	(MCMR-) MCM RES FOUND INC.	
XX		
PI	Masuda I, Halligan BD, McCarty DJ, Ryan LM, Barbieri JT, Haas AL;	
DR	WPI: 2000-022330/02.	
XX		
DR	N-PSDB; AA235852.	
XX		
PT	Porcine nucleotide pyrophosphohydrolase coding sequences and protein	
XX		
PS	Disclosure; Column 19-24; 25pp; English.	
XX		

The present sequence represents nucleotide pyrophosphohydrolase (NTPPH) isolated from a Porcine chondrocyte cDNA library. NTPPH catalyses the hydrolysis of nucleoside triphosphates, such as ATP, to a nucleoside monophosphate and PPI. This is an important factor in the generation of PPI, which then forms calcium pyrophosphate dihydrate (CPPD). CPPD



CC crystal deposition in human articular cartilage is a common age-dependent  
 CC event often associated with degeneration of articular tissues and/or with  
 CC acute attacks of arthritis. Knowledge of the enzymes involved in this  
 CC process is important for combating, e.g. arthritis. Sufficient amounts  
 CC of the enzyme would also be useful for allowing crystallisation to  
 CC determine the enzyme structure. It can also be used as an antigen to  
 CC raise antibodies useful in diagnosis as a marker of joint tissue  
 CC breakdown and of vesicle formation within articular tissues. The cDNA  
 CC can also be used in in situ hybridisation to determine the presence and  
 CC location of mRNA for the parent molecule within cartilage and other  
 CC tissues. The cDNA of NPPH can be used to generate the protein for  
 CC use in screening methods for modulators of the enzyme.

Sequence 599 AA;

Query Match 45.9%; Score 2898.5; DB 21; Length 599;  
 Best Local Similarity 91.1%; Pred. No. 3.8e-254;  
 Matches 544; Conservative 23; Mismatches 29; Indels 1; Gaps 1;

OY 588 PLGEVVGEDPMALFLPSSRFKONGEPKVKASVFLPDRNISTATAOTDLNFTND 647  
 DB 4 plgmdggedpmgeleipsksfyrngpeptgkvasvflpdrnistataaqsdlind 63  
 OY 648 ECDTFPLRTYGMFVDFRDEVTSEPLNAGKVHLSDSTQVKMPEHISTYKLSLNDPTGL 707  
 DB 64 ecdtflrtlygmfvdfdeaaeslnvqkvxhldstqvkmpelvmkmlvslndpdtl 123  
 OY 708 WEEEDGFKEFNGRNRKREDRFLVGNLEIRERRLFNLDVPSRRCFVKVRAYSERFLPS 767  
 DB 124 weeedgfkesgrnrkredrflvgnleirerlfnldvpsrrcfikvrayserflps 183  
 OY 768 EIOGVVSVITLLEPRTGLSNPRAMGRFDSYITGNACVAFDDSDPAKSAVYLAS 827  
 DB 184 eqigvsvvitlleprtglsnpramgrfdsyitgnacvafddsdpaaksavylas 243  
 OY 828 LAGEELQAVESPPKNPNAIGVQPYLNKLNRDHEPRYKTAFOISMAKPPNSAE 887  
 DB 244 lageelqavessppknnpnagvpylnklnrddhepryktafcoddsdpaysaylas 303  
 OY 888 ESNGRPIYAEFLRACEBAPPSAAHFRFYQIEGDRYDNTVPNEDDPMSTEDYLAWMPK 947  
 DB 304 esngripiyaeplracedapppsaahfrfyqiegdrydntvpneddpmstedylawmpk 363  
 OY 948 PHEFACVYKYIVGPLENVNVRSMNGHRRRTGKLYGIRVRSRRDDQNNVSAQLE 1007  
 DB 364 phefacykyivgplenvvnvrsmngghrrrtgklygirvrsrrddqnnvsasacle 423  
 OY 1008 FKCSGMLYDQDVRDRLVIVIPQSCRRASVNPMLHEYLVNHLPLAVNNDTSEYTLAPL 1067  
 DB 424 fkcsgmlydqdvrdrtlkvivpgschrassvnmplheylvnhlplavnnntseytclapl 483  
 OY 1068 DPLGHNNGIYVTQDPTAKEIALGRCFDGTSDGSSSRIMKSNVGVALTFCNVEROVGRQ 1127  
 DB 484 dplghnngiyvtqdptakeialgrcfdsdgsrvmkshnvgtaltfcnverovgqg 543  
 OY 1128 SRFQYLGSRPASPAGYVQGVNPSRQOARSGORGSGVVASLRFVVAQOPLIN 1184  
 DB 544 srfqylgstrpaspagytgvnpsrroarsgorgsgvvaslrfvvaqoplsh 599

RESULT 6

AA49502 standard; Protein: 890 AA.

AA49502;

10-JAN-2000 (first entry)

Human METH2 protein.

Human; METH1; METH2; anti-angiogenic; metalloprotease thrombospondin;  
 cancer; diagnosis; hyperproliferative disorder; autoimmune disease;

KW angiogenesis inhibitor; abnormal wound healing; inflammation;  
 KW rheumatoid arthritis; psoriasis; endometrial bleeding disorder;  
 KW diabetic retinopathy; macula degeneration; haemangioma; detection;  
 KW arterial-venous malformation; immune deficiency.

OS Homo sapiens.

PN W09937660-A1.

PD 29-JUL-1999.

PF 22-JAN-1999; 99WO-US01313.

PR 23-JAN-1998; 98US-0072298.

PR 28-AUG-1998; 98US-0098539.

PA (IRUE/) IRUELA-ARISPE L.

PA (HAST/) HASTINGS G A.

PA (RUBE/) RUBEN S M.

PI Irue-la-Arispe L, Hastings GA, Ruben SM.

DR WPI; 1999-590684/50.

DR N-PSDB; AA232001.

PT New isolated metalloprotease thrombospondin polypeptides, useful for  
 PT treating hyperproliferative disorders, cancers or autoimmune disorders

PS Claim 10; Fig 2; 457bp; English.

CC AA232000 and AA232001 encode, and AA49501 and AA49502 represent, human  
 CC metalloprotease thrombospondin (METH) proteins METH1 and METH2  
 CC respectively. METH1 and METH2 have been found to be potent inhibitors of  
 CC angiogenesis both in vitro and in vivo. They can be used for treating  
 CC cancer and other disorders related to angiogenesis including abnormal  
 CC wound healing, inflammation, rheumatoid arthritis, psoriasis,  
 CC endometrial bleeding disorders, diabetic retinopathy, some forms of  
 CC macula degeneration, haemangiomas, and arterial-venous malformations.  
 CC They may be useful in treating deficiencies or disorders of the immune  
 CC system, by activating or inhibiting the proliferation, differentiation,  
 CC or mobilisation (chemotaxis) of immune cells. The etiology of these  
 CC immune deficiencies or disorders may be genetic, somatic, such as  
 CC cancer or some autoimmune disorders, acquired (e.g. by chemotherapy or  
 CC toxins), or infectious. They can also be used to treat inflammatory or  
 CC conditions, both chronic and acute conditions. The products can also be  
 CC used for detection and diagnosis. AA232002 to AA232080, and AA49503 to  
 CC AA49511 represent sequences given in the exemplification of the present  
 CC invention.

Sequence 890 AA;

Query Match 2.7%; Score 170; DB 20; Length 890;  
 Best Local Similarity 21.4%; Pred. No. 7.3e-06;  
 Matches 110; Conservative 49; Mismatches 224; Indels 132; Gaps 25;

OY 50 LESPGEWTLTFNIDYPGGKDYERLDAIRFYGD--RVCARPLREARTTWPAGSTGQ 107  
 DB 427 ldpagaaapl-ptlgyrmaalyldqgcrqlfpglfrhpnatsaqdvcqlwchdgaep 485  
 OY 108 VHGSPREFGFLNDEQRGQNCNMYVFLCPGSLKRDTRT-----WSPSPMSK 160  
 DB 486 lchtkngslpwadgpcpgphlcsegs---clp---eeveerphrvydgwagwpawge 538  
 OY 161 CSAACGQNGVOTRTRICLAEMWVSLCSEASESGOHCMG-----ODCTACDLTCPMQGVNAD 215  
 DB 539 csrtcg-gvgvqisheckd-----pepgngyryclgrrakyschtee--cpqd-gks 587  
 OY 216 CDACMCODFMLGAVSLPGAPASGAATYLLTKRPLKLTQTDSDGRFRIPGLC-PDGKSI 274  
 DB 588 fregqcekyuaynycdmdgn-----llqwpkkygvgvprd---icklctararise 635

```

QY 275 LKTRKVFAPIVLTMPTSLKAATIKAEFVAETPYVMNPE--TKARRAGQSVSLCCCKA 332
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 636 fkvfeakvidgtlclgpet--laicvrgcvkagcdhvdspkrlidkcgvcgkgnscrkv 693
QY 333 TGRPRDPKIFYMND-----TLDD-----PSLYKHESKLVLRKLOHQHAGEYFCKA 378
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 694 sgsiltplny--gyndivtlipagatnidvkgrshpvgndgnylalktad---gqylilng 747
QY 379 QSDAGAVKS-----KVQLIVTASDETPPCNPVPESEYILRL---PHDCF----- 418
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 748 nlalsalegdllvkgtllkysgsiatlerlqsfirpdpelvtqlltvpgevfppkvtylf 807
QY 419 -----QANRNSFY-----DVGRCPVKTCAGQODNGIRCRD---A 450
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 808 fvpndvdfsmgskerratnllgpllnagwylgdwsecscstcgagwgrttvecrdpsga 867
QY 451 VQNCGSGISKTEFERIQCSTYLPRKVAKEGSCQRC 485
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 868 satcnkalkped-----akpcesqlic 888

```

## RESULT 7

ID AAB50003 standard; Protein: 890 AA.

AC AAB50003:

DT 19-MAR-2001 (first entry)

DE Human METH2.

Human; METH2; metalloproteinase; thrombospondin; angiogenesis inhibition;  
 cancer therapy; benign tumour; ocular angiogenic disease;  
 rheumatoid arthritis; psoriasis; wound healing; endometriosis;  
 vasculogenesis; granulation; hypertrophic scar; nonunion fracture;  
 scleroderma; trachoma; vascular adhesion; myocardial angiogenesis;  
 coronary collateral; cerebral collateral; arteriovenous malformation;  
 ischaemic limb angiogenesis; Osler-Weber syndrome; plaque  
 angioblastoma; fibromuscular dysplasia; haemophilic joint;  
 Crohn's disease; atherosclerosis; birth control.

OS Homo sapiens.

PN WO200071577-A1.

PD 30-NOV-2000.

PE 25-MAY-2000; 2000MO-US14462.

XX 25-MAY-1999; 9905-0318208.  
 PR 20-JUL-1999; 9905-0144882.  
 PR 10-AUG-1999; 9905-0147823.  
 PR 13-AUG-1999; 9905-0373658.  
 PR 22-DEC-1999; 9905-0171503.  
 PR 22-FEB-2000; 2000US-0163792.

PA (HUMA-) HUMAN GENOME SCI INC.

PA (SMIK-) SMITHKLINE BEECHAM CORP.

PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

PA (TRUE-) IRUELA-ARISPE L.

PA (HAST-) HASTINGS G A.

PA (RUBE-) RUBEN S M.

PA (JONAK-) JONAK Z L.

PA (TRULL-) TRULLI S H.

PA (FORN-) FORMAN J A.

PA (TERR-) TERRETT J A.

PI Irueja-Arispe L, Hastings GA, Ruben SM, Jonak ZL, Trulli SH;

PI Fornwald JA, Terrett JA;

XX WPI: 2001-025136/03.

XX N-PSDB; AAC90058.

XX METH1 and METH2 polynucleotides and encoded polypeptides, used to  
 PT inhibit angiogenesis in the treatment of disorders such as cancer,  
 PT rheumatoid arthritis and psoriasis -  
 PS Claim 15; Fig 2; 768pp; English.

CC The present sequence is human METH2 (ME for metalloproteinase and TH for  
 CC thrombospondin). METH2 can be used for inhibiting angiogenesis in an  
 CC individual, and for treating cancer, benign tumours, an ocular angiogenic  
 CC disease, rheumatoid arthritis, psoriasis, delayed wound healing,  
 CC endometriosis, vasculogenesis, granulations, hypertrophic scars, nonunion  
 CC fractures, scleroderma, trachoma, vascular adhesions, myocardial  
 CC angiogenesis, coronary collateral, cerebral collaterals, arteriovenous  
 CC malformations, ischaemic limb angiogenesis, Osler-Weber syndrome, plaque  
 CC neovascularisation, telangiectasia, haemophilic joints, angiodysplasia,  
 CC fibromuscular dysplasia, wound granulation, Crohn's disease or  
 CC atherosclerosis. METH2 can also be used in birth control. METH2 can also  
 CC be used in diagnostic methods for the prognosis of cancer.

SQ Sequence 890 AA;

Query Match 2.7%; Score 170; DB 22; Length 890;  
 Best Local Similarity 21.4%; Pred. No. 7.3e-06;  
 Matches 110; Conservative 49; Mismatches 224; Indels 132; Gaps 25;

```

QY 50 LSPGEMTLEFNIDYPGKGDERLDAIRFYGD--RVCARPLRLERTDWTAGSTGQ 107
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 427 ldpagaaapl-ptglpymalagldqgqrqtlfpdftncpntsaqdcvaqlwchtdaep 485
QY 108 VVHSPREGFWCLNRBQRPQGNCSNYVRFCLPSPSLRPTERT-----WSPFWMSK 160
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 486 lchknsglpmadgtppgphlcsesg---clp---eeerpkpvydgwmapwmgwe 538
QY 161 CSAACGOTGVYTRIRICLAENVSLCSASEGQCMG-----QDCTACDLTPMGQVNAD 215
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 539 csrcrg--gyvdfshreckd-----pepqngryclgrakysqchee--cpdp--gks 587
QY 216 CDACMCDPMFHGAVSLPGAPASGAIAVLLTKPRKLLTQDSGRRIRPOLC--PDGKSI 274
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 588 fregqcekyanayntdamgn-----llqwpkyayvsprd---rcklfcargrse 635
QY 275 LKTRKVFAPIVLTMPTSLKAATIKAEFVAETPYVMNPE--TKARRAGQSVSLCCCKA 332
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 636 fkvfeakvidgtlclgpet--laicvrgcvkagcdhvdspkrlidkcgvcgkgnscrkv 693
QY 333 TGRPRDPKIFYMND-----TLDD-----PSLYKHESKLVLRKLOHQHAGEYFCKA 378
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 694 sgsiltplny--gyndivtlipagatnidvkgrshpvgndgnylalktad---gqylilng 747
QY 379 QSDAGAVKS-----KVQLIVTASDETPPCNPVPESEYILRL---PHDCF----- 418
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 748 nlalsalegdllvkgtllkysgsiatlerlqsfirpdpelvtqlltvpgevfppkvtylf 807
QY 419 -----QANRNSFY-----DVGRCPVKTCAGQODNGIRCRD---A 450
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 808 fvpndvdfsmgskerratnllgpllnagwylgdwsecscstcgagwgrttvecrdpsga 867
QY 451 VQNCGSGISKTEFERIQCSTYLPRKVAKEGSCQRC 485
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 868 satcnkalkped-----akpcesqlic 888

```

## RESULT 8

ID AAY04145 standard; Protein: 481 AA.

AC AAY04145:

DT 15-JUN-1999 (first entry)

DE Rat Tango-76 protein.

```

XX Rat; Tango-71; Tango-73; Tango-74; Tango-76; Tango-83; diagnosis;
KW detection.
XX Rattus sp.
OS WO9907850-A1.
XX 18-FEB-1999.
PD 18-FEB-1999.
XX 06-AUG-1998; 98WO-US16502.
PF 06-AUG-1998; 98WO-US16502.
XX 05-SEP-1997; 97US-0058108.
PR 06-AUG-1997; 97US-0054966.
XX (MILL-) MILLENNIUM BIOTHERAPEUTICS INC.
XX Goodearl ADJ, Holtzman DA;
PI WPI: 1999-167426/14.
XX N-PSDB; AAX19958.
DR New TANGO polypeptides and nucleic acids encoding them - useful as
PT diagnostic agents and for treating disorders caused by aberrant
PT expression of TANGO
XX Claim 8; Fig 6; 84pp: English.
PS The present sequence represents rat Tango-71. Tango polypeptides are
XX useful for identifying compounds which bind the polypeptide via direct
CC binding, competition binding assays or Tango-71, -73, -74, 76 or -83-
CC mediated signal transduction. Tango polypeptides are also useful for
CC identifying modulating compounds by determining effect on Tango activity.
CC Tango polypeptides and nucleic acids are useful for diagnosing diseases
CC related to aberrant expression of Tango, and Tango polypeptides are
CC useful for raising antibodies which can be used in diagnostic assays for
CC detection of Tango, and also for generating anti-idiotypic antibodies for
CC prevention and protection.
XX Sequence 481 AA;
SQ
Query Match 2.7%; Score 167.5; DB 20; Length 481;
Best Local Similarity 20.1%; Pred. No. 4.2e-06;
Matches 97; Conservative 46; Mismatches 156; Indels 183; Gaps 24;
OY 118 WCLNRE-GRPGNCSNVTYRF-----LCPPGS-LRRDTER-----IMSPWP 157
DB 67 wcrhdseplchtkngslwadygpcgphclidgscvltreenpkavvdgdkvpgp 126
OY 158 WSKGSAAG-----QTGVOTRTRICLAEMWSLCSSEASEGCHCMGDCCTACTDIT 206
DB 127 wgcgrtcgsgyqifnrecdnpapnggrfcigerky-----gsckteecp----- 173
OY 207 CPMGVNADCAOCAGODMMLHGAVSLPGGAPRSGAATYILTRPKLTLQOTSDGRF---- 262
DB 174 -pungsfte---gqcekyay-----nhtldignflqvw 203
OY 263 -RIPSLCP-----DGKSLIKTKVKFAPVILMTPKTSKAKATKAIEVRAETPYM 311
DB 204 pkysgvsprrdcklcrargsefkvfdglcpgdt--lalcvrgqcvkagcdh 261
OY 312 VIANPE--TKARAGOSVSLCCATGKPRPRDKYFWTHND-----TLID-----PSLY 355
DB 262 vnsprkklckgcvgcgtactkvsqsftrfisy--gyndivlipagatnldvkqsrhpyq 319
OY 356 KHESLTVLRKTLQHOAGGEFCKAQSDAGAVKSKV----- 389
DB 320 ndgsylalktan-----gqyllngalaiaieqdlmkgtllkysgmatlerlgsfgalp 375
OY 390 ----AQLIVTASDETPCNPVPE--SYLIRLPHD-----CFRONATNSFYV----- 427
DB 376 eplvtvglltvsgevfpr----pkvkytftfvpndtfnvgskskerastnllqslpyewwlg 431

```

```

OY 428 DVGRCPVKTCAGQDNGTRCRD----AVQNCGISKTEREILQSGSYTLPTVAKESCSQ 483
DB 432 dwsecptcggwqrrtvecrldpsqasdtcdalkped-----akpcgsq 477
OY 484 RC 485
DB 478 pc 479
RESULT 9
AAB72300
ID AAB72300 standard; Protein; 1072 AA.
AC AAB72300;
DE 14-MAY-2001 (first entry)
XX Human ADAMTS-10 alternative amino acid sequence.
DE ADAMTS-N; disintegrin; metalloprotease; thrombospondin type I motif;
KW tumour cachexia; inflammation; dermatosparaxis; EDS-VIIC; angiogenesis;
KW Ehlers-Danlos syndrome type VIIC; articular cartilage erosion; human;
KW metastasis; embryogenesis; egg implantation; ADAMTS-10.
XX Homo sapiens.
XX WO20011074-A2.
XX 15-FEB-2001.
XX 03-AUG-2000; 2000WO-US21223.
XX 06-AUG-1999; 99US-0369364.
XX (CLEV-) CLEVELAND CLINIC FOUND.
PA (APTE/) APTE S S.
PA (HURS/) HURSKAINEN T L.
PA (HIRO/) HIROHATA S.
XX Apte SS, Hurskainen TL, Hirohata S;
XX WPI: 2001-159978/16.
DR N-PSDB; AAF63448.
XX Murine and human 'A Disintegrin-like And Metalloprotease domain with
PT Thrombospondin type I motifs' proteins and the nucleic acids encoding
PT them, useful for treating e.g. tumours, inflammation and arthritis -
PS Disclosure: Fig 16; 181pp: English.
XX This invention relates to murine and human ADAMTS-N (A disintegrin-like
CC and metalloprotease domain with thrombospondin type I motifs) proteins,
CC designated ADAMTS-5, 6, 7, 8, 9, 10 and R1. Also included in the
CC invention are cDNA sequences encoding the proteins, and antibodies
CC specific for the proteins. The nucleic acid sequences and proteins may be
CC used in the prevention, diagnosis and treatment of diseases associated
CC with inappropriate ADAMTS-N expression. Disorders that may be treated
CC using the nucleic acids, proteins and antibodies include, for example
CC tumour cachexia, inflammation, dermatosparaxis in cattle or Ehlers-Danlos
CC syndrome type VIIC (EDS-VIIC) in humans, erosion of articular cartilage
CC in arthritic (both inflammatory and non-inflammatory) disease,
CC angiogenesis, tumour growth and metastases, and they may also be used for
CC controlling embryogenesis and implantation of fertilised eggs. The
CC present sequence represents human ADAMTS-10.
XX Sequence 1072 AA;
SQ
Query Match 2.6%; Score 161.5; DB 22; Length 1072;
Best Local Similarity 20.6%; Pred. No. 6e-05;
Matches 100; Conservative 62; Mismatches 178; Indels 145; Gaps 26;

```

```

QY 107 QVHGPSREGFMCINROR-----PGQMCNNTV-----RFLCPGSLRRDTER 150
    | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 458 gckygwcseicwclskscnrcitnsipaegtlcqtldkwcgkrcvcpfspegvd 517
    | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 151 IWSFSPWMSCSACGGTQVOTRRICLAENVSLCSSESGCHCMQ-----DCTACDL 205
    | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 518 awgawcpwgdstrfcg-gvyssssrhcdsprtl-----gqkyclgertrhsctldd- 569
    | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 206 TCPMGQVNADACMCDPMLHGAVSLP-----GGAPPA-----S 239
    | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 570 -cpgp--sqdfrvqgsef-----dsiprfgkfkwtkyrgyggvkaclsiaegfnfyt 621
    | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 240 GAATYLLTKTP-----KLITQDSGRRRIPGLCPDGKSLMTK 279
    | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 622 eraaavdgfpcrpdtdvcsgeckhvgcdrvlgslredkcrvcg--gqgsacetleg 679
    | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 280 VKFAPL-----VLMPKTSLKATIKAEFVRAETRPYWMNETARRAGGSVSLCK 331
    | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 680 v-lspaspgagvedvvpilpysvh---lfiqdlnlslshalk-----gqgesllle 727
    | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 332 A-TGKPRPDYFYWHNDTLDPRLSKHESKLVLRKLOQHAGYEYFCKAOSDAGAVKSKVA 390
    | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 728 glpgrtp-----hrlplagttfqlrgpdqvgsllea-lpplnasli 768
    | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 391 QLIYASDETPCNVPESYLIRLPHDCFOQNTNSFYI-DVGRCPVKTCAGQDNGIRCRD 449
    | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 769 -vmulartelpra---lryfnapiardslppyswbyapwlcksaagcaqsqvagvecrn 823
    | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 450 AVON-----CCGISKTEEREIQSGSYTLPTKVA---KCC--SCQRCEPSTRITV-RGR 495
    | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 824 qldsavaphycasahsklprktraentleppdwvvgvnwslcstrcdagaytrsvvcqtr 883
    | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 496 VSAAD 500
    | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 884 vsaae 888

RESULT 10
AAB72284
ID AAB72284 standard; Protein; 905 AA.
XX
AC AAB72284;
XX
DT 14-MAY-2001 (first entry)
XX
DE Murine ADAMTS-8 amino acid sequence.
XX
KW ADAMTS-N; disintegrin; metalloprotease; thrombospondin type I motif;
KW tumour cachexia; inflammation; dermatosparaxis; EDS-VIIC; angiogenesis;
KW Ehlers-Danlos syndrome type VIIC; articular cartilage erosion; mouse;
KW metastasis; embryogenesis; egg implantation; ADAMTS-8.
XX
OS Mus musculus.
XX
PN WO20011074-A2.
XX
PD 15-FEB-2001.
XX
PF 03-AUG-2000; 2000MO-US21223.
XX
PR 06-AUG-1999; 99US-0369364.
XX
PA (CLEV-) CLEVELAND CLINIC FOUND.
PA (APTE/) APTE S S.
PA (HURS/) HURSKAINEN T L.
PA (HIRO/) HIROHATA S.
XX
PI Apte SS, Hurskainen TL, Hirohata S;
DR WPI; 2001-159978/16.
DR N-PSDB; AAF63441.
XX
PT Murine and human 'A Disintegrin-like and Metalloprotease domain with

```

```

PT Thrombospondin type I motifs' proteins and the nucleic acids encoding
PT them, useful for treating e.g. tumours, inflammation and arthritis -
XX
PS Claim 1; Fig 5; 181pp; English.
XX
CC This invention relates to murine and human ADAMTS-N (A disintegrin-like
CC and metalloprotease domain with thrombospondin type I motifs) proteins,
CC designated ADAMTS-5, 6, 7, 8, 9, 10 and R1. Also included in the
CC invention are cDNA sequences encoding the proteins, and antibodies
CC specific for the proteins. The nucleic acid sequences and proteins may be
CC used in the prevention, diagnosis and treatment of diseases associated
CC with inappropriate ADAMTS-N expression. Disorders that may be treated
CC using the nucleic acids, proteins and antibodies include, for example
CC tumour cachexia, inflammation, dermatosparaxis in cattle or Ehlers-Danlos
CC syndrome type VIIC (EDS-VIIC) in humans, erosion of articular cartilage
CC in arthritis (both inflammatory and non-inflammatory) disease,
CC angiogenesis, tumour growth and metastases, and they may also be used for
CC controlling embryogenesis and implantation of fertilised eggs. The
CC present sequence represents murine ADAMTS-8.
XX
SQ Sequence 905 AA;

Query Match 2.5%; Score 160; DB 22; Length 905;
Best Local Similarity 20.6%; Pred. No. 6; Le-05;
Matches 109; Conservative 48; Mismatches 183; Indels 190; Gaps 27;

QY 65 PGKGYERLDIAIRFYVG-----DRVCARPLREARTDMPAGSTGVVHGPS 113
    | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 455 pghstlyelqgckqifgpdfirhcnptsvedic---vqlcarhd-----sdepichkn 506
    | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 114 REGFWCLNRDQRPQNCNNTVREFLCPGSS--LRDTER-----IWSFSPWMSKCSAA 164
    | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 507 glllwadgtpcpgph-----lclsgscvllkedvnpkavdgdgwrgwmpgqsrst 557
    | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 165 CGGTGVTQTRTRICLAENVSLCSSESGCHCMG-----QCTACDLTCPMGVNADCDAC 219
    | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 558 cg-ggldfsmrecdnpn-----punggrfclgerlvkyscntee--cpnngksfregqc 608
    | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 220 MCDPFLHGAVSLPGAPASGAAYLLTKTPKILTOFDSGRF-----RIPGLCP----- 269
    | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 609 ekynaynh-----tdlgnflgwprkysvgsprdrck 640
    | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 270 -----DCKSLIKTKVKFAPILVLTMPKTSLKATIKAEFVRAETPYWMNPE--TKARRA 322
    | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 641 lforargsefkvfeakvldgltcgpdt--lslcvrgcvkagcdhvnspkklckgycv 698
    | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 323 GQSVSLCKKAWGKRPDPKXYFWYHND-----TLID-----PSLYKHESKLVLRKLOQ 368
    | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 699 gqkylactrkisgltfisy--gynldvltlpagatnldvkqshpynndgsyjalktan- 755
    | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 369 HQAGEYFCKAQSDAGAVKSKV-----AQLIVTASD 398
    | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 756 ---gqyllnglnlalsalegdlvkgtllkysgsmatlerlqsfalpepiltvqlltvsge 812
    | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 399 ETPCNPVPE-SYLIRLPHDCFOQNTNS-----FYDVGRCPVKTCAGQ 440
    | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 813 vfp---pkvryltftypndmfsvgnskeratnllgslpsaewlgdwecp--storgs 867
    | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 441 -QDNGIRCRD---AYQNCGISKTEEREIQSGSYTLPTVAKESCQRC 485
    | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 868 wgrttvecrdpsggsatcdcaalkped-----akpcgsqpc 903
    | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 11
AAB72288
ID AAB72288 standard; Protein; 1081 AA.
XX
AC AAB72288;
XX
DT 14-MAY-2001 (first entry)
XX
DE Human ADAMTS-10 amino acid sequence.

```

XX ADAMTS-N; disintegrin; metalloprotease; thrombospondin type I motif;  
 KW tumour cachexia; inflammation; dermatosparaxis; EDS-VIIC; angiogenesis;  
 KW Ehlers-Danlos syndrome type VIIC; articular cartilage erosion; human;  
 KW metastasis; embryogenesis; egg implantation; ADAMTS-10.  
 XX Homo sapiens.  
 OS  
 XX  
 XX MO20011074-A2.  
 XX  
 XX 15-FEB-2001.  
 XX  
 XX 03-AUG-2000; 2000WO-US212223.  
 XX  
 XX 06-AUG-1999; 99US-0369364.  
 XX  
 XX (CLEV-) CLEVELAND CLINIC FOUND.  
 XX (APPE/) APPE S S.  
 XX (HORS/) HORSKAINEN T L.  
 XX (HIRO/) HIROHATA S.  
 XX  
 XX Apte SS, Hurskainen TL, Hirohata S;  
 XX  
 XX WPI; 2001-159978/16.  
 XX N-PSDB; AAF63445.  
 XX  
 XX Murine and human 'A Disintegrin-like And Metalloprotease domain with  
 PT Thrombospondin type I motifs' proteins and the nucleic acids encoding  
 PT them, useful for treating e.g. tumours, inflammation and arthritis -  
 XX  
 PS Claim 1; Fig 9; 181pp; English.  
 XX  
 CC This invention relates to murine and human ADAMTS-N (A disintegrin-like  
 CC and metalloprotease domain with thrombospondin type I motifs) proteins,  
 CC designated ADAMTS-5, 6, 7, 8, 9, 10 and R1. Also included in the  
 CC invention are cDNA sequences encoding the proteins, and antibodies  
 CC specific for the proteins. The nucleic acid sequences and proteins may be  
 CC used in the prevention, diagnosis and treatment of diseases associated  
 CC with inappropriate ADAMTS-N expression. Disorders that may be treated  
 CC using the nucleic acids, proteins and antibodies include, for example  
 CC tumour cachexia, inflammation, dermatosparaxis in cattle or Ehlers-Danlos  
 CC syndrome type VIIC (EDS-VIIC) in humans, erosion of articular cartilage  
 CC in arthritic (both inflammatory and non-inflammatory) disease,  
 CC angiogenesis, tumour growth and metastases, and they may also be used for  
 CC controlling embryogenesis and implantation of fertilised eggs. The  
 CC present sequence represents human ADAMTS-10.  
 CC  
 SO Sequence 1081 AA;

Query Match 2.5%; Score 157.5; DB 22; Length 1081;  
 Best Local Similarity 20.2%; Pred. No. 0.00014;  
 Matches 98; Conservative 62; Mismatches 180; Indels 145; Gaps 25;

OY 107 QVHGSPREGFGLNREOR-----PGQNCNVTV-----RFLCPPGSLRDPER 150  
 DB 466 qckyeveveselwclsknrcinsipaaegtlcqhltldkycwkrvcyfrfsgvpaq 525  
 OY 151 IWSFSPWMSKCSAAGCGVQTRTRICLAEMVSLCSEASEGQHCQMGQ-----DCTACDL 205  
 DB 526 awgptwtpgdcstcrg-gvsvsssrhcdsprtli-----ggkylgerrrrhnsntcd- 577  
 OY 206 TCFPMQVNAADCACMGQDFMLHGANVSLP-----GGAA-----S 239  
 DB 578 -cpbg--sqdfrevqcaef-----dsipfrgkyfwktyrggvkacsltslaegfnfyt 629  
 OY 240 GAATVILRTTP-----KLTFQTDSDGFRRIJGLCPDCKSLIKITK 279  
 DB 630 eraaavogtprcptvdlcvsgeckhvgcatvigsdlredkcrvcg--gdsaacetleg 687  
 OY 280 VKFAPL-----VLTFPKTSLKATIAKFAEFVRAETPYVMVNPETKARRAQSVSLCK 331  
 DB 688 v-fspaspagaygedvfwlpkgsvh---ffqdlnlslnhlaik-----gdgsllle 735

OY 332 A-TGKPRDKYFWYHNDLFLDPSLYKHESKLVLRKLOOHQAGEYFCRAQSDGAVKSVYA 390  
 DB 736 glpypcpq-----hrlplagttffqlrqpdpqvslea-lypinaali 776  
 OY 391 OLIVTASDETPCNVPESYLRLPHDFOMATNSFYY-DVGRCPVKTCAGQDNGIRCRD 449  
 DB 777 -vmvlarlelpa---lryfnaplardslppyswhypawtkcsaqgsgvqavecrn 831  
 OY 450 AVQN-----CCGISKEERETQCSGYTLPRK--VAKECSQRCFTE-----TRSTVGR 495  
 DB 832 qlssavaphycashsklpkracntecppdvwvgnwslcsrscdagvistsvcqr 891  
 OY 496 VSAAD 500  
 DB 892 vsaae 896

RESULT 12  
 AAB00042  
 ID AAB00042 standard; Protein: 1152 AA.  
 XX  
 XX AAB00042;  
 AC  
 XX  
 DT 08-NOV-2000 (first entry)  
 XX  
 DE Human thrombospondon-1 (TSP-1).  
 XX  
 KW TSP-1; TSP-2; COMP; cartilage oligomeric matrix protein;  
 KW thrombospondin; angiogenesis; tumour; treatment; cancer;  
 KW arthritis; psoriasis; diabetic retinopathy; corneal graft rejection;  
 KW glioma.  
 KW  
 XX Homo sapiens.  
 OS  
 XX  
 XX Key Location/Qualifiers  
 FT Region 361..416  
 FT /label= Type 1 repeat region  
 FT Region 417..473  
 FT /label= Type 1 repeat region  
 FT Region 474..530  
 FT /label= Type 1 repeat region  
 XX  
 PM WO200044908-A2.  
 XX  
 PD 03-AUG-2000.  
 XX  
 PF 01-FEB-2000; 2000WO-US02482.  
 XX  
 PR 01-FEB-1999; 99US-0118053.  
 XX  
 PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.  
 PI Lawler JW;  
 XX  
 XX WPI; 2000-514823/46.  
 XX  
 CC New nucleic acids are described which encode a protein comprising  
 CC the second and third type 1 repeats of human TSP (thrombospondin)-1,  
 CC but not the TGF (transforming growth factor)-beta activation region  
 CC of human TSP-1. The nucleic acid of TSP (thrombospondin)-1 containing  
 CC the second and third type-1 repeats and the COMP (cartilage  
 CC oligomeric matrix protein) assembly sequence (COMP/TSP-1) was  
 CC produced by PCR (polymerase chain reaction). Expression of COMP/TSP-1  
 CC caused inhibition of the growth of tumours in mice models.  
 CC Thus the nucleic acids and proteins may be useful for treating  
 CC angiogenesis related diseases such as cancer (by reducing the rate of



```

Db 596 padact-nhm-----gehrce-ntdpgynclpcpprftgsq 628
OY 384 ---AVKSKVAOLIVTASDETPCNPVPESYLIRLPHDCFONATNSF---YVD-VGRCPVK 435
Db 629 pf9gvgvshatankayc-----kprnptcdg-----thdcnkakcnylghysdpmrycck 679
OY 436 TCAGGQDNGIRC 447
Db 680 --pgyagngllc 689

RESULT 14
AAB21251
ID AAB21251 standard; Protein: 680 AA.
XX
AC AAB21251;
XX
DT 23-FEB-2001 (first entry)
XX
DE Human metalloproteinase ADAMTS-2.
XX
KW Human; ADAMTS2; metalloproteinase; ADAM;
KW a disintegrin and metalloproteinase domain; thrombospondin domain;
KW vaccine; neurotropic; neuroprotective; antiparkinsonian;
KW cerebroprotective; cytosolic; antiarthritic; immunosuppressive;
KW Alzheimer's disease; Parkinson's disease; stroke; cancer; arthritis;
KW autoimmune disease; brain tumour; brain injury.
XX
OS Homo sapiens.
XX
PN WO200053774-A2.
XX
PD 14-SEP-2000.
XX
PF 08-MAR-2000; 2000WO-US06237.
XX
PR 08-MAR-1999; 99US-0264585.
XX
PA (NEUR-) NEUROCRINE BIOSCIENCES INC.
XX
PI Kelnner GS, Clark M, Maki RA;
XX
DR WPI: 2000-594326/56.
DR N-PSDB: AAA95821.
XX
PT Polynucleotide encoding novel members of a disintegrin,
PT metalloproteinase and thrombospondin domain protein family used to
PT prevent and treat Alzheimer's disease, cancer and autoimmune diseases -
XX
PS Claim 12; Fig 2; 129pp; English.
XX
CC The present sequence is human metalloproteinase ADAMTS-2. The
CC ADAMTS family of proteins is closely related to the ADAM (A Disintegrin
CC and Metalloproteinase Domain) family. Members of the ADAMTS family
CC contain a thrombospondin domain in addition to the disintegrin and
CC metalloproteinase domains found in the ADAMs. ADAMTS polypeptides are
CC useful for the manufacture of medicaments for treating conditions
CC associated with neuroinflammation and/or neurodegeneration, such as
CC Alzheimer's disease, Parkinson's disease and stroke. They are also
CC useful for treating conditions associated with cell proliferation, cell
CC migration, inflammation and/or angiogenesis, such as cancer, arthritis
CC and autoimmune diseases. They can be used to treat patients afflicted
CC with an invasive tumour, a brain tumour or brain injury.
XX
SQ Sequence 680 AA:

```

```

Query Match 2.5%; Score 156; DB 21; Length 680;
Best Local Similarity 21.2%; Pred. No. 8.5e-05;
Matches 97; Conservative 45; Mismatches 181; Indels 134; Gaps 23;
OY 118 WCLNREQRGRGSCSNVTVRFL---CPGSL-----RRTERTI-----WSPWSPW 158

```

```

Db 267 wchtdgaeplichtknsgslpwadgtipcpgnhlcsesclpceeverpkpyvdgvgawpawp 326
OY 159 SKCSACGOTGVQVOTRTRICLAEWVSLCSSEASEGHCNMG-----ODCTACDILTCEMGQVN 213
Db 327 gescrtcg-gyvgqfshreckd-----pepqngryclgrtakyschtee--cpdd--g 375
OY 214 ADCDACMKODEFHLGAVSLPGGAPASGAATYLLTTPKLLTQTSDDGRFRIPGLC-PPGK 272
Db 376 ksfreqqcekyanayltmdgn-----llgwbpkyagvsprd---rcklforargr 423
OY 273 SILKITKVFAPITVDMPTSLKAAPTIFAEFVRAETPYVMNM--PETKARRAGQSVSLCC 330
Db 424 sefkvfeakvldgtlcspect--laicvrgcvkagcdhvdsfwkldkcygcgkygnsct 481
OY 331 KATGKRPDPKYFWYHND-----TLTD-----PSLYKHESKLVLRKLOHQHGEYFC 376
Db 482 ksgsgsltpnny--gynldvltlpagatnldvkqshpnyqndgnylalktad----gyll 535
OY 377 KASDAGAVKS-----KVAQLIVTASDETPCNPVPESYLIRL----- 413
Db 536 ngulaaisaiedqllvkgtllkysgsiatlerlqsfripdltvgllavpgevfpkvky 595
OY 414 ----PHD-----CFONATNSFY-----DVGRCPVKTCAGGQDNGIRCRD--- 449
Db 596 ttfvprndvdfsmgskeratnltqpllhaqvvlgdwsecscstcgawqrrtvecrdpsg 655
OY 450 -AVQNCGGSISKTEEREDICSGYTLPTKYAKKESCCORC 485
Db 656 gasatcnkalkped-----akpcesqlc 678

RESULT 15
AAM81030
ID AAM81030 standard; Protein: 1496 AA.
XX
AC AAM81030;
XX
DT 10-MAY-1999 (first entry)
XX
DE Melanoma associated antigen MC50.
XX
KW MC50; melanoma gene-50; melanoma associated antigen; human;
KW T cell epitope; cancer; lung cancer; rhabdomyosarcoma; diagnosis;
KW therapy; vaccine.
XX
OS Homo sapiens.
XX
PN WO9855133-A1.
XX
PD 10-DEC-1998.
XX
PF 04-JUN-1998; 98WO-US11533.
XX
PR 06-JUN-1997; 97US-0870941.
XX
PA (REGC ) UNIV CALIFORNIA.
PA (UWSC-) UNIV SOUTHERN CALIFORNIA.
XX
PI Deans RJ, Kan-Mitchell J, Minev BR, Mitchell MS;
XX
DR WPI: 1999-080820/07.
DR N-PSDB: AAV99922.
XX
PT New MC50 melanoma associated antigen fragments - used to develop
PT products for the detection, treatment and prevention of
PT MC50-expressing cancers, e.g. melanoma, lung cancer or
PT rhabdomyosarcoma
XX
PS Claim 1; Page 45-49; 79pp; English.
XX
CC This polypeptide comprises a portion of a new human melanoma
CC associated antigen, designated MC50. The amino acid sequence was

```





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 9, 2001, 11:49:00 ; Search time 36.04 Seconds  
(without alignments)  
4346.536 Million cell updates/sec

Title: US-09-609-383-2  
Perfect score: 6317  
Sequence: 1 MVGTAKWVSFLVLEVTSL.....QSGVASTLRPRVAQOPLIN 1184

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues  
Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: SPREMBL\_16:\*  
2: sp.archaea:\*  
3: sp.bacteria:\*  
4: sp.fungi:\*  
5: sp.human:\*  
6: sp.invertebrate:\*  
7: sp.mammal:\*  
8: sp.mhnc:\*  
9: sp.organelle:\*  
10: sp.phage:\*  
11: sp.plant:\*  
12: sp.podent:\*  
13: sp.spodent:\*  
14: sp.virus:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6317	100.0	1184	4 075339	075339 homo sapien
2	2898.5	45.9	599	6 019112	019112 sus scrofa
3	168.5	2.7	1948	4 013332	013332 homo sapien
4	164.5	2.6	1637	6 09XSV8	09XSV8 bos taurus
5	161.5	2.6	1077	4 09H324	09H324 homo sapien
6	155.5	2.5	7962	4 010465	010465 homo sapien
7	155	2.5	1496	4 092626	092626 homo sapien
8	154	2.4	779	5 087136	087136 manduca sex
9	154	2.4	7107	5 09V477	09V477 drosophila
10	152	2.4	1250	11 088971	088971 mus muscula
11	152	2.4	6658	5 076281	076281 drosophila
12	150	2.4	4123	4 075851	075851 homo sapien
13	147.5	2.3	788	5 022631	022631 caenorhabdi
14	147	2.3	1444	5 017591	017591 caenorhabdi
15	146	2.3	1239	5 061541	061541 drosophila
16	146	2.3	1239	5 09V3X0	09V3X0 drosophila
17	146	2.3	1302	5 061542	061542 drosophila
18	145.5	2.3	4488	5 09TXK4	09TXK4 caenorhabdi
19	144	2.3	1099	11 P97527	P97527 rattus norv

## ALIGNMENTS

RESULT 1	075339	PRELIMINARY;	PRT; 1184 AA.	
AC	075339;			
DT	01-NOV-1998 (TRENBLREL. 08, Created)			
DT	01-NOV-1998 (TRENBLREL. 08, Last sequence update)			
DT	01-MAR-2001 (TRENBLREL. 16, Last annotation update)			
DE	CARTILAGE INTERMEDIATE LAYER PROTEIN.			
GN	CLIP.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=ARTICULAR CARTILAGE;			
RX	MEDLINE=98389785; PubMed=9722584;			
RA	Lorenzo P., Neame P., Sommarin Y., Heinegard D.;			
RT	"Cloning and deduced amino acid sequence of a novel cartilage protein (CLIP) identifies a proform including a nucleotide			
RT	pyrophosphohydrolase.";			
RL	J. Biol. Chem. 273:23469-23475(1998).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RA	Nakamura I., Okawa A., Ikegawa S., Takaoka K., Nakamura Y.;			
RT	"Genomic organization, mapping, and polymorphisms of the gene encoding			
RT	human cartilage intermediate layer protein (CLIP).";			
RL	Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RA	Lorenzo P., Aman P., Sommarin Y., Heinegard D.;			
RT	"Pro-CLIP: Gene structure and chromosomal localization.";			
RL	Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.			
CC	-I- SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX			
CC	DOMAIN.			
DR	EMBL; AF035430; AAC3838.1; -			035888 rattus norv
DR	EMBL; AB022430; BAA76692.1; -			09H3V5 homo sapien
DR	EMBL; AF035455; AAF14689.1; -			002661 bos taurus
DR	EMBL; AF035448; AAF14689.1; JOINED.			08TYK4 caenorhabdi
DR	EMBL; AF035449; AAF14689.1; JOINED.			09V3G1 drosophila
DR	EMBL; AF035451; AAF14689.1; JOINED.			076065 homo sapien
DR	EMBL; AF035453; AAF14689.1; JOINED.			099552 homo sapien
DR	EMBL; AF035453; AAF14689.1; JOINED.			09S62 homo sapien
DR	EMBL; AF035453; AAF14689.1; JOINED.			035158 rattus norv
DR	EMBL; AF035453; AAF14689.1; JOINED.			001761 caenorhabdi
DR	EMBL; AF035453; AAF14689.1; JOINED.			09UJPI homo sapien
DR	EMBL; AF035453; AAF14689.1; JOINED.			018263 caenorhabdi
DR	EMBL; AF035453; AAF14689.1; JOINED.			09XTD5 caenorhabdi
DR	EMBL; AF035453; AAF14689.1; JOINED.			094780 homo sapien
DR	EMBL; AF035453; AAF14689.1; JOINED.			094779 homo sapien
DR	EMBL; AF035453; AAF14689.1; JOINED.			09IAR9 gallus gall
DR	EMBL; AF035453; AAF14689.1; JOINED.			097394 drosophila
DR	EMBL; AF035453; AAF14689.1; JOINED.			098918 gallus gall
DR	EMBL; AF035453; AAF14689.1; JOINED.			09IAR9 gallus gall
DR	EMBL; AF035453; AAF14689.1; JOINED.			014887 homo sapien
DR	EMBL; AF035453; AAF14689.1; JOINED.			009993 caenorhabdi
DR	EMBL; AF035453; AAF14689.1; JOINED.			09XTD2 caenorhabdi
DR	EMBL; AF035453; AAF14689.1; JOINED.			09QW24 rattus sp.
DR	EMBL; AF035453; AAF14689.1; JOINED.			09ULX4 homo sapien
DR	EMBL; AF035453; AAF14689.1; JOINED.			09UDY8 homo sapien
DR	EMBL; AF035453; AAF14689.1; JOINED.			09VZ24 drosophila

DR InterPro: IPR000884; -  
 DR InterPro: IPR001451; -  
 DR InterPro: IPR002086; -  
 DR InterPro: IPR003006; -  
 DR Pfam: PF00047; 1g; 1.  
 DR Pfam: PF00090; tsp.1; 1.  
 DR PROSITE: PS00070; ALDEHYDE\_DEHYDR\_CYS; UNKNOWN.1.  
 DR PROSITE: PS00101; HEXAPEP\_TRANSFERASES; UNKNOWN.1.  
 DR SMART: SM00209; TSP1; 1.  
 SQ SEQUENCE 1184 AA; 132538 MW; 4449F0537CC99C3 CRC64;

Query Match 100.0%; Score 6317; DB 4; Length 1184;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1184; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGTKAWVFSLVLEVTSLGROTMLOSRYVOPGKKNPSIFAKPADTLESPEGEMTLF 60  
 1 MGTKAWVFSLVLEVTSLGROTMLOSRYVOPGKKNPSIFAKPADTLESPEGEMTLF 60  
 DB 1 MGTKAWVFSLVLEVTSLGROTMLOSRYVOPGKKNPSIFAKPADTLESPEGEMTLF 60  
 QY 61 NIDYPGKGYERLDAIRFYGGDRCARPLRLAARTTDMTPAGSTGOVHGSPEGEFWCL 120  
 61 NIDYPGKGYERLDAIRFYGGDRCARPLRLAARTTDMTPAGSTGOVHGSPEGEFWCL 120  
 DB 61 NIDYPGKGYERLDAIRFYGGDRCARPLRLAARTTDMTPAGSTGOVHGSPEGEFWCL 120  
 QY 121 NREORPGONCNTVVRFLCPGSLRDRTERLWSPWSPKCSAACGOTGVOTRTRICLAE 180  
 121 NREORPGONCNTVVRFLCPGSLRDRTERLWSPWSPKCSAACGOTGVOTRTRICLAE 180  
 DB 121 NREORPGONCNTVVRFLCPGSLRDRTERLWSPWSPKCSAACGOTGVOTRTRICLAE 180  
 QY 181 MVSICSEASEGOGHOMGODTACDLTCPMGOVNADCDACMOPMLHGAASLPGAPASG 240  
 181 MVSICSEASEGOGHOMGODTACDLTCPMGOVNADCDACMOPMLHGAASLPGAPASG 240  
 DB 181 MVSICSEASEGOGHOMGODTACDLTCPMGOVNADCDACMOPMLHGAASLPGAPASG 240  
 QY 241 AATYLLTKTKLLTQDSDGRFRPLGICLPGKSLTKTKKFAPIYVLTMRKTSIKAAITK 300  
 241 AATYLLTKTKLLTQDSDGRFRPLGICLPGKSLTKTKKFAPIYVLTMRKTSIKAAITK 300  
 DB 241 AATYLLTKTKLLTQDSDGRFRPLGICLPGKSLTKTKKFAPIYVLTMRKTSIKAAITK 300  
 QY 301 AEFVRAETPYMNMNPETKARAGOSVSLCCKATGKPRPDYFWFHNHTLLDPSLYKHESK 360  
 301 AEFVRAETPYMNMNPETKARAGOSVSLCCKATGKPRPDYFWFHNHTLLDPSLYKHESK 360  
 DB 301 AEFVRAETPYMNMNPETKARAGOSVSLCCKATGKPRPDYFWFHNHTLLDPSLYKHESK 360  
 QY 361 LVLRKLOOHAGGEYFCKAQSADAGAVSKVAQLVYASDETPCNPVPESYLLRLPHDFQN 420  
 361 LVLRKLOOHAGGEYFCKAQSADAGAVSKVAQLVYASDETPCNPVPESYLLRLPHDFQN 420  
 DB 361 LVLRKLOOHAGGEYFCKAQSADAGAVSKVAQLVYASDETPCNPVPESYLLRLPHDFQN 420  
 QY 421 ATNSFYDVGRCPYKTCAGQDNGIRCRDAVONCCGISKEBERIQSGTTLPTKVAKEC 480  
 421 ATNSFYDVGRCPYKTCAGQDNGIRCRDAVONCCGISKEBERIQSGTTLPTKVAKEC 480  
 DB 421 ATNSFYDVGRCPYKTCAGQDNGIRCRDAVONCCGISKEBERIQSGTTLPTKVAKEC 480  
 QY 481 SCORCTETRSIVGRVSAADNGEPMRGHYMGNRSVMTGYGTFTLHVPDTERLYLT 540  
 481 SCORCTETRSIVGRVSAADNGEPMRGHYMGNRSVMTGYGTFTLHVPDTERLYLT 540  
 DB 481 SCORCTETRSIVGRVSAADNGEPMRGHYMGNRSVMTGYGTFTLHVPDTERLYLT 540  
 QY 541 FVDRLOKFVNTTKVLPENKGSADFHEIKMLRKEPITLLEAMETNIIPLGEVVGEDPMAE 600  
 541 FVDRLOKFVNTTKVLPENKGSADFHEIKMLRKEPITLLEAMETNIIPLGEVVGEDPMAE 600  
 DB 541 FVDRLOKFVNTTKVLPENKGSADFHEIKMLRKEPITLLEAMETNIIPLGEVVGEDPMAE 600  
 QY 601 LEIPSRSEFYQNGEPYIGKVKASVTEFLDPNNISTATAOQDLNFINDEGCTFLPTKGMF 660  
 601 LEIPSRSEFYQNGEPYIGKVKASVTEFLDPNNISTATAOQDLNFINDEGCTFLPTKGMF 660  
 DB 601 LEIPSRSEFYQNGEPYIGKVKASVTEFLDPNNISTATAOQDLNFINDEGCTFLPTKGMF 660  
 QY 661 SVFPRDEVTESEPLNAGVKVHLNSTOVKMPENHISTYKLMSLNPDGTGWEESGDFKFNOR 720  
 661 SVFPRDEVTESEPLNAGVKVHLNSTOVKMPENHISTYKLMSLNPDGTGWEESGDFKFNOR 720  
 DB 661 SVFPRDEVTESEPLNAGVKVHLNSTOVKMPENHISTYKLMSLNPDGTGWEESGDFKFNOR 720  
 QY 721 RNKREDRTFLVNLREIRRLFLNLDVPSRRCVKAAYSESEFLPSEQIQGVIVSYNL 780  
 721 RNKREDRTFLVNLREIRRLFLNLDVPSRRCVKAAYSESEFLPSEQIQGVIVSYNL 780  
 DB 721 RNKREDRTFLVNLREIRRLFLNLDVPSRRCVKAAYSESEFLPSEQIQGVIVSYNL 780  
 QY 781 EPRTGFLSNRANGRDSVITGNGACVPAFCDDQSDAYSAVYASLAGEELQAVESSP 840  
 781 EPRTGFLSNRANGRDSVITGNGACVPAFCDDQSDAYSAVYASLAGEELQAVESSP 840  
 DB 781 EPRTGFLSNRANGRDSVITGNGACVPAFCDDQSDAYSAVYASLAGEELQAVESSP 840  
 QY 841 KFRPNAIGVQPYLNLKNTYRTDHEPDRVKKTAFOISMAMPNRSASENSGPIYAFENLR 900

DB 841 KFRPNAIGVQPYLNLKNTYRTDHEPDRVKKTAFOISMAMPNRSASENSGPIYAFENLR 900  
 QY 901 ACEAPPSAAHFRFYOLEGRDYDNTVPEFNEDDPMSTEDYLAMWPKPMFEFRACYIKVKI 960  
 901 ACEAPPSAAHFRFYOLEGRDYDNTVPEFNEDDPMSTEDYLAMWPKPMFEFRACYIKVKI 960  
 DB 901 ACEAPPSAAHFRFYOLEGRDYDNTVPEFNEDDPMSTEDYLAMWPKPMFEFRACYIKVKI 960  
 QY 961 VGPLEVNVSRNMGGTHTRTYGLKYGIRDYRSTREDRQDPNVSAACLEFKSGMLYQDDRV 1020  
 961 VGPLEVNVSRNMGGTHTRTYGLKYGIRDYRSTREDRQDPNVSAACLEFKSGMLYQDDRV 1020  
 DB 961 VGPLEVNVSRNMGGTHTRTYGLKYGIRDYRSTREDRQDPNVSAACLEFKSGMLYQDDRV 1020  
 QY 1021 DRTLKYIIPGSCRRASVNMLEHYLNLPLAVNNDTSEYTMALPLDPLGHNYGYITVT 1080  
 1021 DRTLKYIIPGSCRRASVNMLEHYLNLPLAVNNDTSEYTMALPLDPLGHNYGYITVT 1080  
 DB 1021 DRTLKYIIPGSCRRASVNMLEHYLNLPLAVNNDTSEYTMALPLDPLGHNYGYITVT 1080  
 QY 1081 DDPRTAKEIALGRCFEDTSDGSSRIKSNVGVALLFENCYERVOGRSAPQYQSTPAQS 1140  
 1081 DDPRTAKEIALGRCFEDTSDGSSRIKSNVGVALLFENCYERVOGRSAPQYQSTPAQS 1140  
 DB 1081 DDPRTAKEIALGRCFEDTSDGSSRIKSNVGVALLFENCYERVOGRSAPQYQSTPAQS 1140  
 QY 1141 PAAGTVGGRVPSRRQGRASRGOSGVASLRPPRYAQQPLIN 1184  
 1141 PAAGTVGGRVPSRRQGRASRGOSGVASLRPPRYAQQPLIN 1184  
 DB 1141 PAAGTVGGRVPSRRQGRASRGOSGVASLRPPRYAQQPLIN 1184

RESULT 2  
 019112 PRELIMINARY; PRT; 599 AA.  
 AC 019112;  
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
 DT 01-NOV-1998 (TrEMBLrel. 05, last sequence update)  
 DT 01-NOV-1998 (TrEMBLrel. 08, last annotation update)  
 DE NUCLEOTIDE PYROPHOSPHOHYDROLASE (FRAGMENT).  
 OS Sus scrofa (Pig).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
 NCBI\_TaxID=9823;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95164702; PubMed=7860751;  
 RA Masuda I., Hamada J., Haas A.L., Ryan L.M., McCarty D.J.;  
 RT "A unique ectonucleotide pyrophosphohydrolase associated with porcine chondrocyte-derived vesicles."  
 RT J. Clin. Invest. 95:699-704(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97473522; PubMed=9332376;  
 RA Masuda I., Halligan B.D., Barbieri J.T., Haas A.L., Ryan L.M., McCarty D.J.;  
 RT "Molecular cloning and expression of a porcine chondrocyte nucleotide pyrophosphohydrolase."  
 RT Gene 197:277-287(1997).  
 DR EMBL; U63114; AAC48770.1; -  
 KW Hydrolase.  
 FT NON TER  
 SQ SEQUENCE 599 AA; 67432 MW; CCC4D08B8C1CF2D1 CRC64;

Query Match 45.9%; Score 2898.5; DB 6; Length 599;  
 Best Local Similarity 91.1%; Pred. No. 5e-221;  
 Matches 544; Conservative 23; Mismatches 29; Indels 1; Gaps 1;  
 QY 588 PLCEVVGEDPMAELEIPSRSEFYQNGEPYIGKVKASVTEFLDPNNISTATAOQDLNFINDEGCTFLPTKGMF 660  
 588 PLCEVVGEDPMAELEIPSRSEFYQNGEPYIGKVKASVTEFLDPNNISTATAOQDLNFINDEGCTFLPTKGMF 660  
 DB 588 PLCEVVGEDPMAELEIPSRSEFYQNGEPYIGKVKASVTEFLDPNNISTATAOQDLNFINDEGCTFLPTKGMF 660  
 QY 648 EGDPTFLRTYGMFSVPRDEVTESEPLNAGVKVHLNSTOVKMPENHISTYKLMSLNPDGTG 707  
 648 EGDPTFLRTYGMFSVPRDEVTESEPLNAGVKVHLNSTOVKMPENHISTYKLMSLNPDGTG 707  
 DB 648 EGDPTFLRTYGMFSVPRDEVTESEPLNAGVKVHLNSTOVKMPENHISTYKLMSLNPDGTG 707  
 QY 708 WEESGDFKFNORRNKREDRTFLVNLREIRRLFLNLDVPSRRCVKAAYSESEFLPSEQIQGVIVSYNL 780  
 708 WEESGDFKFNORRNKREDRTFLVNLREIRRLFLNLDVPSRRCVKAAYSESEFLPSEQIQGVIVSYNL 780  
 DB 708 WEESGDFKFNORRNKREDRTFLVNLREIRRLFLNLDVPSRRCVKAAYSESEFLPSEQIQGVIVSYNL 780  
 QY 781 EPRTGFLSNRANGRDSVITGNGACVPAFCDDQSDAYSAVYASLAGEELQAVESSP 840  
 781 EPRTGFLSNRANGRDSVITGNGACVPAFCDDQSDAYSAVYASLAGEELQAVESSP 840  
 DB 781 EPRTGFLSNRANGRDSVITGNGACVPAFCDDQSDAYSAVYASLAGEELQAVESSP 840  
 QY 841 KFRPNAIGVQPYLNLKNTYRTDHEPDRVKKTAFOISMAMPNRSASENSGPIYAFENLR 900  
 841 KFRPNAIGVQPYLNLKNTYRTDHEPDRVKKTAFOISMAMPNRSASENSGPIYAFENLR 900  
 DB 841 KFRPNAIGVQPYLNLKNTYRTDHEPDRVKKTAFOISMAMPNRSASENSGPIYAFENLR 900



SO	SEQUENCE	1948 AA;	217080 MW;	7DC049EC03171136 CRC64;
FT	VASPLIC	1350	1365	MISSING (IN PTSP-MED).
FT	VARSPIC	1366	1366	S -> G (IN PTSP-MED).
FT	CONFLICT	309	310	YT -> YHP (IN REF. 2).
FT	CONFLICT	428	429	SA -> RP (IN REF. 2).
FT	CONFLICT	742	745	LGCV -> RSPA (IN REF. 2).
FT	CONFLICT	765	773	GABRGPR -> RREGRGRS (IN REF. 2).
FT	CONFLICT	910	910	R -> P (IN REF. 2).
FT	CONFLICT	986	995	AAEGGANAV -> GRISRRRTL (IN REF. 2).
FT	CONFLICT	1195	1196	TEV -> SL (IN REF. 2).
FT	CONFLICT	1431	1431	F -> S (IN REF. 2).
FT	CONFLICT	1546	1546	E -> D (IN REF. 4).
FT	CONFLICT	1587	1587	V -> A (IN REF. 4).
FT	CONFLICT	1705	1705	N -> K (IN REF. 2).

Query Match	2.7%;	Score 168.5;	DB 4,	Length 1948;
Best Local Similarity	20.2%;	Pred. No. 0.0006;		
Matches 195;	Conservative 117;	Mismatches 366;	Indels 289;	Gaps 48

QY	295	KAATIKAEVAEAPRYAMNETARAGOSJCLCCATCKPRDXKEWYHNPTLDPSTL	354
Dp	121	KLTVLREDOQLPSGEFNIDMGQLKVERTKATIMCAASGNPDE-ITWKKDLFVDP	179
QY	355	YKHESK-----LVLRKLOOQAGEYFCKAOSDGAVKSKYAOLIVTASDE	399
Dp	180	SNGRIKOLRSEFESTPIRGALQIESSEEDDQKECVCATINSAGVRYSSPANLYVRELR	239
QY	400	TPCRNPVESYILRLPHOCFQNA-INSTRYYDGR-CP-VKICAGODNGICRDAVONCCG	456
Dp	240	VR-RVAPRFSTILPESHIMPGNVTICVAVGSPMPVKKMOGAE-----	284
QY	457	ISKTEEREIOCSGYTLTPKVAECCSCORCTREIRIVGRVSADNGEPMRFGHYMNSR	516
Dp	285	--LHPEDDMRGVRLVELIDYKDSANTCYAMSL-GVLEAV-----AQ	325
QY	517	VSMTCY-KGFTLHVPODTERVLVFDRLQKFEVNTTKVLRFNKKSAVHEIKMLRKE	575
Dp	326	ITVKSILPKAAGTPWVTENTATISLITWDSONDPVSYVYIEKSKSODGQYQK----	379
QY	576	PITLEAMETNIPRIGEVGEDPMALFET-----PSSFYQONEPPIYIKYKAS	623
Dp	380	----EDITTRYISG---GSPNSEYELIWMSAVNSIQGPPSESVYTRGE---QAPAS	428
QY	624	VTEFLPNISATQAQTDLNFINEDEGTFFRLRYGMPFSVPDEFVTSSEPLNACKVXNLD	683
Dp	429	A-----PRNVQARMLSATMTMIOUME-----PYERNGLLR-GYRYIYIME-----	467
QY	684	STQYKMEBHISTYKVLMSLNPDTGLMEBEGDFKFNORRNKREDRTFLVGLNLEIRERL	743
Dp	468	-----BEH-----PVGHW-----QKNHVDSDSLITTVGSL-----	491
QY	744	LDPESHRCFVKAYVRSERFLP-SQDI-----OGVIVSYINL--PERTGFLSNPRWG-	794
Dp	492	--LEDETYIVKALAFSVSGDPLSDPIQVKTQOGVGPOMNIRAEARSE-TSITLSWSP	547
QY	795	-REDSVI-----TGPNGACVPAFCDQSDPAYSAVVLASLAGEELQAVESSPKFNPNA	846
Dp	548	PROGSIKIELLREGOHGREV-----GRIFPDITSYVVEDLKNPTETARLARLAK-SPOG	601
QY	847	IGVQPLKLNKLRITDHEDR-----YKKAFOISMAKPRPNSAEESNGPIYAFE--	897
Dp	602	LGATPVVRORTLOSKSPAPPODKCVSVSTAILVSWRPPR-----ETHGALVGSVRY	658
QY	898	-----NLFACEAPPSAHPFYQIE-GDAYDYVTVPFNEDDP-----MSWTEP	940
Dp	659	RPLGSEDPREXVNGIRPTTQILLLEALKEMWQIRITTVAHTEGPRESSPVVVRTDED	718
QY	941	YLAWMPKPM-----FRACIYIKVIKYLEVAVNVRSNMGSTARFTYVGKIGICIDRV--STRD	995
Dp	719	VPSAPPRKVEALNANATIAIVLMGPRP-----GRGNGQLRG--YQVHYIYRMGAEGB	768
QY	996	RDPBNVSAACLERKCSGMLYD-ODRVDRTLVKVIPOGSCRRASVBNPLMEIYVNNHPLAV	105

[illegible]

RESULT 4  
Q9XSV8  
ID Q9XSV8 PRELIMINARY; PRT; 1637 AA

DT	01-NOV-1999	(TREMblrel_12, Created)
DT	01-NOV-1999	(TREMblrel_12, last sequence update)
DT	01-MAR-2001	(TREMblrel_16, last annotation update)
DE	SCO-SPONDIN	(FRAGMENT).
DE	SCO-SPONDIN	.
GN	SCO-SPONDIN	.
OS	Bos taurus	(Bovace).
OC	Eukaryota	(Molozoa); Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Euthetia; Cetartiodactyla; Ruminantia; Pecora; Bovidae;	
OC	Bovidae; Bovinae; Bos.	
OX	NCBI_TaxID=9913;	
OX	NCBI_TaxID=9913;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RA	Gobron S., Creveaux I., Didier R.,	
RT	"Characterization of cattle SCO-Rpind."	
RL	Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.	
DR	EMBL; A1133488; CAB46239.1; -.	
DR	HSP; P56682; ICCV.	
DR	InterPro; IPR000359; -.	
DR	InterPro; IPR000884; -.	
DR	InterPro; IPR001007; -.	
DR	InterPro; IPR001545; -.	
DR	InterPro; IPR002919; -.	
DR	Pfam; PR00007; Cys_knot; 1.	
DR	Pfam; PF00090; tsp_1; 11.	
DR	Pfam; PF01826; TIL; 5.	
DR	PROSITE; PS01225; CTCX_2; 1.	
DR	PROSITE; PS00261; GLYCO_HORMONE_BETA_1; UNKNOWN_2.	
DR	PROSITE; PS01208; VWFC; UNKNOWN_1.	
DR	SMART; SM00041; CT; 1	
FT	NON_TER	
SO	SEQUENCE	1637 AA; 173346 MW; 4C5BABI3D346C925 CRC64;

Query Match	2.6%	Score 164.5	DB: 6	Length 1637
Best Local Similarity	21.6%	Pred. No. 0.00095		
Matches 100, Conservative	26	Mismatches 160	Indels 177	Gaps 22

QY	94	ARTDMP----	AGSTGYVH-GSPRGFCMC--LNRQGRGCMCNATVFELC-----	139
Dd	646	SRSWAMSPCSKSCGPAPAQOSRFRSTSGSWAPECREEOSOCPOSCPPLLOCHRR	7050	
QY	140	-----	PGLSRDTE-----RLWSPPMSKSCSAACGGYGTFRICL	178
Dd	706	SLGSMLTODGCQCSCCTPEEGICEDACAGLTANTPSPMSDCPVSCG-GGNOYFRIRVCY	764	
QY	179	AEMVSLCSEASEECHOHCMDG----	C7ACDLTC--EMGOVNADCDACMCOFWM	225
Dd	765	A-----SAPPRGSSPCILCPDVOSORCGIWPCCALPDPTCSGMGWGPCSRSCG-----	810	
QY	226	LHGAVSLPGCAPASGAAILYLTK-----	TRKLTLQTJSDGRFRIRGLGCLPBGSKSLIKI	277
Dd	811	-----PGLASRASCSOPLLAEEPACNASTSPILDIOA-----CYVG-	846	



```

Db 2688 GPNITTSKGC---NIVTTEKTCILEILN---STPRDAGGY-----SCIEINENAGR----- 2731
QY 216 CDACMCQCFMHLGAVSLGAPASAAIYLLTKTKLLTQDSDGRFRIPGLCPDGKSL 275
Db 2732 -DVC-----GALVSTL----- 2741
QY 276 KITVKKFAPVILMPKTSIKAAITKAEPVRAETPYVMNPETKARAGOSVSLCKATKG 335
Db 2742 -----EPPYFTELEPPLANAAGDSVSLCCOVAGT 2770
QY 336 PRDQKFWYHNDLLDS-----LYHESKLVLRKLOOHAGEYFCKAOSDAGAVSKV 389
Db 2771 PE-ITVSWMYGDKRLRPTPEYRTFTNNVATLVENKYNINDSGEYTCKAENISGTASKT 2829
QY 390 AOLI-----VTASDETPCNVPP-----ESYLIRLPHDCQ- 419
Db 2830 VFRIOERQLPPSFARQKLDIEQVGLFVTLTCLRLNGSAPLIOVCWYRGVLLR-DHENLQT 2888
QY 420 -----MATN-----SFYDVGRCPYKTC 437
Db 2889 SFVDNVAATLKILQTDLSHSGQYSCSASNPLGTASSARLARERPKKSPFEDIKPVSIDVT 2348
QY 438 AGOODNGIRCRDAVONCCGIS-KTEREIOCSG-YTLP-----TKARECSQ-R 484
Db 2949 AGESAD-FECHVTGAOPMRITWSKDNKEIRPGGNITTCVGNTPHLRIKLKVGKDSGQYT 3007
QY 485 CETRSIVGRVSA-----ADNGEPHREHVWNGSRVMTGKGT 525
Db 3008 COATNDVGCMSAQSLVKEPRFKVAKLASKYAKOGESITQLECKTIGSBEIKVSWFRND 3067
QY 526 FILHVQDTERLVLTVDRLQKRVNTKVLPPNK----- 559
Db 3068 SELHESWKYN-----MSFINVALLITINEASAEDSGYICAHNGVGDASGTL 3117
QY 560 --KGSVAFHEIKMLRKKEPTLEAME-TNITPLGEVGVGDPAHELEIPRSFTRONGEPY 616
Db 3118 TVKAPPVF-----TOKSPVGAUKGSDVILLQCEISGTTPPEVWVVKRQKQVR-NSKRF 3169
QY 617 IGKVAASVTEFLDPNISTATAQTDLNFINDEG-DTF-----PLRTYGMS----- 661
Db 3170 KITSHKPTNLHILNLEASDVGEHCATNEGSDTSCSVKFKPPRVKSLDSTILI 3229
QY 662 ---VDFREYVS-EPLNA-----GKVYHLDSTOVKMEHISTVKLMS----- 700
Db 3230 GDAVELRAIVEGFOPISVVWLKDRGEVIRESENTRISFIDNIATLQLSPEASNSGKYIC 3289
QY 701 -LNPDTGLMEEGDKFENQNR--NKREDREFVGN----- 733
Db 3290 QIKNDAGMRECSAVILVEPARIIEKPEPMTVTGPNPALCECVYGTPELSAKWFKDGRE 3349
QY 734 -----LEIRERRLFNLDVPS---RRCVVKVRAVRSERFLPS- 767
Db 3350 LSAOSKHITFINKVASIKIPCAMSDKGLAFEVKNSVGSNCTVSYHV--SDRIYPPS 3407
QY 768 -----EIOIGVIVINLEPRTGFLSNPRANGR--DSVITGPNACVPAFCD----- 813
Db 3408 FIRKLAKVNNALGVSYLECVRSG-SAPISYGMFQDNEIYSGPK--QSSFSNACTLN 3464
QY 814 ----DOSPDAYSAYVLASLAGEE---LQAVESPKF--NPNALGV 849
Db 3465 LSLLEPDTGTITCVANANVAGSDCSAVLTVOEPPSEFQDPDSVEV 3510

```

```

OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BONE MARROW;
RX MEDLINE=97191544; PubMed=9039502;
RA Nagase T., Seki N., Ishikawa K., Ohira M., Kawarabayashi Y., Ohara O.,
RA Tanaka A., Kotani H., Miyajima N., Nomura N.;
RT "Prediction of the coding sequences of unidentified human genes. VI.
RT analysis of cDNA clones from cell line KG-1 and brain.";
RL DNA Res. 3:321-329(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=MELANOMA;
RX MEDLINE=95048383; PubMed=7959781;
RA Weiler S.R., Taylor S.M., Deans R.J., Kan-Mitchell J., Mitchell M.S.,
RA Trent J.M.;
RT "Assignment of a human melanoma associated gene M650 (D2S448) to
RT chromosome 2p25.3 by fluorescence in situ hybridization.";
RL Genomics 22:243-244(1994).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=MELANOMA;
RA Mitchell M.S., Kan-Mitchell J., Minev B., Edman C., Deans R.J.;
RT "Identification of a novel melanoma gene (M650) - likely the gene for
RT IL-1 receptor antagonist - which encodes epitopes recognized by human
RT cytolytic T lymphocytes.";
RL Submitted (Oct-1999) to the EMBL/Genbank/DBJ databases.
CC -1 SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX
CC DOMAIN.
DR EMBL: D86983; BAA13219.1;
DR EMBL: AF200348; AAF06354.1;
DR HSSP: P05164; ICXP.
DR InterPro: IPR000483;
DR InterPro: IPR001007;
DR InterPro: IPR001536;
DR InterPro: IPR001611;
DR InterPro: IPR002007;
DR InterPro: IPR002016;
DR InterPro: IPR003006;
DR Pfam: PF00047; Ig; 4.
DR Pfam: PF00093; wvc; 1.
DR Pfam: PF00141; peroxidase; 1.
DR Pfam: PF00560; LRR; 5.
DR Pfam: PF01463; LRRCT; 1.
DR PRINTS: PR00457; ANPEROXIDASE.
DR PROSITE: PS01208; WPCF; UNKNOWN_1.
DR SMART: SM00214; WVC; 1.
DR NON_TER 1
SQ SEQUENCE 1496 AA; 167209 MW; E9B9A706BFB1ABFF CRC64;

```

Query Match 2.5%; Score 155; DB 4; Length 1496;  
Best Local Similarity 18.5%; Pred. No. 0.0047;  
Matches 186; Conservative 115; Mismatches 352; Indels 352; Gaps 46;

```

QY 297 ATIAEFVRAETPYVMNPETKARAGOSVSLCKATGKRPDKYFWYHND----- 347
Db 251 AITPPEELNCEPRITSEPDADVTSGNTYVFTCRAGGNPKPE-IWLRNNNELSMKTD 309
QY 348 --TLDPSLYKHESKLVLRKLOOHAGEYFCKAOSDAGAVSKVVAOLITVASETP- 401
Db 310 RLNLDD-----DGTLMITQTOETDGTITQCMKKNVAGEVKTQEVLRFGSPARTFV 363
QY 402 -----C-----NPPESYLIR-----LPHDCFQMATNSFYVDGRCPVKT 436
Db 364 QPONTFVLGSEVTLCSATNGHPPTISWTRGDRTPLVPDPKNITPS-----GGLYIGN 418
QY 437 CAGQODNGIRCRDAVONCCGISKT-----EER-----EIOCSGYTLP 473

```

```

Db 419 WV-QDGSGEYACSATNNIDSVHAFIVQALPQVTVTPQDRVIVIEGQVDFQCEAKGNP 477
QY 474 -----TKVAKESGOCRCFTSRISVGRVSA---DNCEMRGHHVYMGSRYSMGYKG 524
Db 478 PVIYANTKGSQSLVDRLHVLISGTLRISGVALHDOGOYECQAVNITGSOKV----- 530
QY 525 TETLH-----VPQDTERLVLFVDFLQKRVNTKTV----- 554
Db 531 --VAHLVQPRVTPVFASIPSDT-----IVEGAVNQVLCCSSOGEPERA 572
QY 555 LPPNKG-----SAVH-----ETIKMRKRPITLMEATNIIPLGEVVGEDPMAAELE 602
Db 573 ITMKNKGVOYTESGKFHISPEGEFLTINDVGPADAGRECVARNTIGSASV---SMVLSVN 629
QY 603 IPRSPYRQWGEPIYG-KVYASVTFIDPRNISTATAQTLNFINDEGDFPLRTGMFS 661
Db 630 VPPVS---RKGDPFVATSTVEALATVDRAINSTRT-----HLFSPRSPDILLALFR 679
QY 662 VDFREYTSBPLNAGVKVHLIDSTQVAMPREHISTVKLMSINPDTGLMEERGDEFKFN--- 718
Db 680 YP-RDPTVQARAGEL---FERTLOLQEHVGHGLMVDLNGTSYHNDLVSPQYLNLIA 735
QY 719 -----QRRKREDRTF-----LVGNLE----- 735
Db 736 NLSGCTAHRREVNNCSMCFHQKYRTHDGTGNNLOHPMMGASLTAFERLKSVEYENGENTP 795
QY 736 --FRRERLEN---LDVPESRCFVKYRAYSERFLPSEQVGVISVINLEPRTGFLSNP 790
Db 796 RGINPHRLVGHALPMPR---LVSTTLIGTEVTVPDQFTHMLMO----- 837
QY 791 RANGRF-----DSVITGPNACVPACDDQ-----SPDAYSAYV---LASLAGEE 832
Db 838 --WGFIDHLDSTVVALSGA---RFSGQHCNSVCSNDPCCSVMIIPRDSARARSARC 892
QY 833 LQAVESPKFNPAIGV-----PQRYLKNL-----YRRTDHEDPVK----- 870
Db 893 MFEVRSSPVGSGMTSLMSVYPREQINQTLSTYIDASNYGSTEHRARSIRDLASHRL 952
QY 871 -----KTAFQISMAKPRPNSAESNGPIYAF---ENLRACEEAPPAHFRFQTE 918
Db 953 LRQGIORSKRPLLPRATGPRTECMDESNIPCFLAGOHANEOLIGLISMHTLMFR-E 1011
QY 919 GDRYDNTVAFNEDPMSTEDYLAWMPKMEPRACYIKVIVGPLEVNVSRN-----M 973
Db 1012 HNRITATELKNP---HMDGTI-----YETKVIYGAELQHTTYQHMLPKIL 1056
QY 974 GGTHTRTVGLYIGRIDVSTRDQDPVNSACLEFKCSGMLYDQDRVDTLVVITDQSGC 1033
Db 1057 GEVGMRTLGEYHGY-----DPOINGIFNAFATAF---RGGHTL----- 1093
QY 1034 RRSVPMLEHYLVNHLPLAVNNDTSEYTLAP-----LDPL 1070
Db 1094 ----VAPLLYRLDENFOPIAQDHLPLHKAFFSFRIVNEGIDPL 1134

```

## RESULT 8

```

ID 097136 PRELIMINARY: PRT: 779 AA.
AC 097136;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, last sequence update)
DE 01-MAR-2001 (TREMBLrel. 16, last annotation update)
OS FASCICLIN II GPI-LINKED ISOFORM.
OC Manduca sexta (Tobacco hawkmoth) (Tobacco hornworm).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
OC Sphingidae; Sphingidae; Sphinginae; Manduca.
OX NCBI_Taxid=7130;
RN [1]
RP SEQUENCE FROM N.A.
RA Wright J.W., Snyder M.A., Combes S., Copenhaver P.F.;
RT Fasciclin II and neuronal migration.
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.

```

```

CC -1 SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX
CC DOMAIN.
DR EMBL: AF103899; AAD17918.1; -.
DR InterPro: IPR001777; -.
DR InterPro: IPR003006; -.
DR InterPro: IPR003598; -.
DR Pfam: PF00041; fn3; 2.
DR Pfam: PF00047; fn3; 5.
DR SMART: SM00408; IGC2; 1.
SQ SEQUENCE 779 AA; 86493 MW; BD9600E4DE02BC5 CRC64;

```

Query Match 2.4%; Score 154; DB 5; Length 779;  
 Best local similarity 20.3%; Pred. No. 0.0021;  
 Matches 132; Conservative 88; Mismatches 203; Indels 228; Gaps 35;

```

QY 294 LKAATTKAEFVRAETPYVMNPETKAR-----RAQSVSLCKATGKPRPDYFYVHND 347
Db 211 LABRNKILEVYTA-----PEMERESFVEIKEESAAITCKARKGP-PPRYTMIKAS 261
QY 348 TLDDP-----SLYKHSKLVLRKLOHQAGEFPCAQSDAG-----AVKSKVAQ 391
Db 262 TRENLATTSRFVSNEISGLTFDVRVAGDYGYKICSAVNNAGQNETEIEVEVLVKRIFIE 321
QY 392 LIYASDET-----PC-----NPVPESLRLPH-DFQONATNSFYDYDGRCPVTCAGQ 441
Db 322 LKNTTAPQIDEGLECKATGRPARISFKLSNDRFLNPN---DGGTITETSRQT 377
QY 442 DNGIRCDAYQNCGISKTEEREIQSGYTLPTKVAKEGSCORCTETRSIVRGVSAADN 501
Db 378 GDMESTGVI-TISTLNRTDGLYEC-----VAENDG 408
QY 502 GEPMRGCHV-----YMGNSRYSMTGYKGTFTLHVPODTERLVLFVDR 544
Db 409 GEARRNGHLVEERKPSFHEMPVPIWANNQPNLS---CIAESIPIATIKKRFRELD 464
QY 545 LD-----KPVNTT-----KVLDPNKKGSAPVHEIKMLRKSPITLLEAM 582
Db 465 VETHMLKIRSGPISITITITPLDQDEGVYKCIATNTHGAE-HILOFRAYRPGAVVQA 523
QY 583 ETNIIPLGEV---VGEDPMAELEIPSRSP---YRONGEPIYKVASVTFDPRNIS- 633
Db 524 KOELITATSVTFGIVG-PAEEMGPIPLATYAQYKENG------FDMWLAQN 568
QY 634 -TATAQTLNFINDEGDTPLRTYGMFSYDF----- 664
Db 569 RTWSANSNSLYVEN-----LRP--MFTYDFPFAANQVGVGAMGSPLTIVIMPRSPPE 620
QY 665 ----RDVTSSEPLNACKVKVHLIDSTQVKMPEHISTVKLMSLNDTG---LMEE----- 711
Db 621 QPKRKEDISESLHGN---YADRYELK-----MKVPRHNBEPDLYETQCPVL 667
QY 712 ---GDFKFNORNRKRED-RTPLVGNLEIRERRLFNLDPESRRCVKAAYSERF-LP 766
Db 668 KVSQDMRVADESLCVKEQIESFEVINVEYGLK-----PDT-RYMHILRAHVHGLSP 720
QY 767 SEQIQGVVISVNLDEPRTGFLSN--PRAMGRD---SVITGPNACVPAPFC 812
Db 721 AOLYVOTALGEVNSE-----SNEVPROPFGYDVTASEALKPSSAITTFC 765

```

## RESULT 9

```

ID 09VAF7 PRELIMINARY: PRT: 7107 AA.
AC 09VAF7;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 16, last sequence update)
DE 01-MAR-2001 (TREMBLrel. 16, last annotation update)
OS BT GENE PRODUCT.
GN BT OR CG1479.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

```

OC Ephydroidea; Drosophilidae; Drosophila.  
 RX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champagne M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Mikos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Baller R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,  
 RA Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,  
 RA Jatalin D., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Laso P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mallet B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paule J.M.,  
 RA Palatzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Relbert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spler E., Spreading A.C., Stapleton M., Strong R., Sun E.,  
 RA Svrtkask R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RA "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195(2000).  
 CC -1- SIMILARITY: TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 CC -1- SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX  
 CC DOMAIN.  
 DR EMBL: AE003843; AAF59316.2; -;  
 DR HSSP: P40189; 180U  
 DR Flybase; FBgn0005666; bt.  
 DR InterPro; IPR000255; -;  
 DR InterPro; IPR000577; -;  
 DR InterPro; IPR000719; -;  
 DR InterPro; IPR001777; -;  
 DR InterPro; IPR002290; -;  
 DR InterPro; IPR003006; -;  
 DR Pfam; PF00041; fn3; 39.  
 DR Pfam; PF00047; 19; 20.  
 DR Pfam; PF00069; pkinase; 1.  
 DR PRINTS: PR00014; ENTPEIT1.  
 DR PROSITE; PS00933; FG\_Y KINASES\_1; UNKNOWN\_1.  
 DR PROSITE; PS00012; PHOSPHOTANETHEINE; UNKNOWN\_1.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; UNKNOWN\_1.  
 DR PROSITE; PS00111; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
 KM ATP-binding; Repeat: Serine/threonine-protein kinase; Transferase.  
 SO SEQUENCE 7107 AA; 794479 MW; CC852ABC29ADBFBC CRC64;

Query Match 2.4%; Score 154; DB 5; Length 7107;  
 Best Local Similarity 19.3%; Pred. No. 0.061;  
 Matches 160; Conservative 101; Mismatches 257; Indels 310; Gaps 39;

QY 234 GGAPASGAII-----YLLTKPKLLTQTD-----SDGRFRIPGLCPDGK---- 272  
 Db 2165 GGASITGYIVERKDPNTGKMQKALETSTPDCKARVNDLIAGNKYQFRIMAVNKGKSKPS 2224  
 QY 273 --SIKITKVKVFAPIVLTMPK---TSLKAATIKAEFAVEFPVVMNPETARRAGOSVS 327  
 Db 2225 EPSDQMTKDRFAP-----PRIDRTINIDITIK-----AGHIR 2258  
 QY 328 LCKATGKPRDKYFWYNDITL--DPSLY-----KHESKLVLRKLOOHQGEYFAQS 380  
 Db 2259 FDIKVSQGP-FATKWLILNKARLENDSDNYIMDESFTKLTIVISKRFHSGKTTLAEN 2317  
 QY 381 DAGAVKSVVAOLITVASDEPCNPVPSYILRLPHDFQNAFNSFYDGR--CPVKTC 438  
 Db 2318 ESGRDEASFEVITLD-----KGPPEGPRLVT-----DVHREGCKLWMA 2357  
 QY 439 GQDDGICRCRAVQVCCISKTEEREIOCSYTLPTKAKESQO--KTERRSIVRGVS 497  
 Db 2358 PLDDGGLPTIDHYI-----TEKMDVE--SGRWLPSSGRFKESFALNLEPSHEYKRFVL 2408  
 QY 498 AAD---NCEPMRFGHVYMGNSRVSWTGYKGTFTLVPODTERLVLTVDRLOKEFVNTKV 554  
 Db 2409 AVNTGESEPEPLGEGSVLAKNPFDDEPKRGF-----AYDMKDHVDLWVR 2455  
 QY 555 LPFKKGSVAVHEIMLRKKEPTIT--LEAMETNIIPLGE-----VGEDPMAELEI- 603  
 Db 2456 PPIINDGSPIDGYV--VERREKGTDMKIKGTETITPCLEGCKATVPLTNCEVEFFVK 2513  
 QY 604 -----PSKSEFRONGEPIYIGKVASVFLDPENISTATAQTDLNF--INDEGDTF 652  
 Db 2514 AINAGPGEPSDA-----SKPIITPKRLAKRIDKNIRITNFKSGEPIFLDINSQPA 2568  
 QY 653 PLRT-----YMFVSVERDE 667  
 Db 2569 PVTWNQNNKSVQTSFSHIENLPYNTKYINNPERKDTGLYKISAHNFYQDDVEFOIN 2628  
 QY 668 VTSEPLNNGK-----VKVHLDSTQVAMP-----EHISVTKMSINPTGLMEEGD 713  
 Db 2629 IITKP--GKEPGLVESEVHDKGLKMKKPKKDDGEPVESVLEKFPDPTGLMPLVG- 2684  
 QY 714 FFEQORRRKRDRFVLGNLEIRERRFLNDLPESRCFVKVRAVRSERFLPSQIOGV 773  
 Db 2685 -----RSDGPE-----YVNDGLVPGHDYK-F 2704  
 QY 774 VISYINLEPRIGTSLNPRAMGFDSVITGPNACYPACDDOSPATSAIVLASIGSEL 833  
 Db 2705 RYKAVNKE-----GESEPLET-----LGSIIADP 2729  
 QY 834 GAVESSPKFNPAIGVPOPYLNLKLVYRRTHDEDPVKKKTAPQISMARPPNSAESNGPI 893  
 Db 2730 FSVPTKP-----GVPEP-----TDW-----TANKVELAMPDP--ASDGSPI 2764  
 QY 894 --YAEENL-----RACEAPSAHFREFYQIEGDRDYNTVPNE 931  
 Db 2765 QGYIYEVKDKYSLPWEKALETNSPTPTATVQGLIGNEGYQFRVALNK 2812  
 RESULT 10  
 ID 088971 PRELIMINARY; PRT; 1250 AA.  
 AC 088971:  
 DT 01-NOV-1998 (TREMBLrel. 08. Created)  
 DT 01-NOV-1998 (TREMBLrel. 08. Last sequence update)  
 DT 01-MAR-2001 (TREMBLrel. 16. Last annotation update)  
 DE CDO.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 RX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Kang J.S., Mulieri P.J., Miller C., Sassoon D.A., Krauss R.S.;



```

RT "CDO, a Robo-related cell surface protein that mediates myogenic
RT differentiation."
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX
CC DOMAIN.
DR EMBL: AF090866; AAC43031.1; -.
DR InterPro: IPR001777; -.
DR InterPro: IPR003006; -.
DR InterPro: IPR003598; -.
DR Pfam: PF00041; f03; 3.
DR Pfam: PF00047; f03; 3.
DR SMART: SM00408; Igc2; 1.
DR SEQUENCE 1250 AA; 135362 MW; 7F9AE93D94D8CCF CRC64;

Query Match      2.4%; Score 152; DB 11; Length 1250;
Best Local Similarity 19.5%; Pred. No. 0.061;
Matches 208; Conservative 132; Mismatches 378; Indels 346; Gaps 54;

QY 309 PYWMNPEFKARAGOSVSLCCATGKPRDPKYEYHNDTLDPPLK---HESKLVLRK 365
DB 28 PFVISEPLSAVOKLGRPVYLHCSA--KPYTARIIMLHNGKRLDRNTEQIKIHGTLTILS 85
QY 366 LOOHQAGEYFCKAOSDAGAVKSKVAQL-----IVTASDETP---CNPVE 407
DB 86 LNPISLGGCYCAVANSVGAIVGSPATVSAALGDFDSSIMHVTAEKMTGFIGCR-VPE 144
QY 408 S-----YLIR---LPHDCFONATNSFYDYGRCPVTKCAGQODNGICRDVAVGCCGI 457
DB 145 SNPKAEVRKIKGKMLKH-----STGNYILLPSGNIQVLNVSSDKGSYKC--AAVN---- 194
QY 458 SKTEEREICSGYTL-----PTKVAKEC--SCORCETRS 490
DB 195 PVTSELAKEVPTGRKLLVSPSSNGPHILPALSOALAVLPHSVTLIECVSGVPASQVW 254
QY 491 IYRGVSAADNGEPMKFGHYVGN-----SRVSMYKGTFTLVHPDTERLVLEVD 543
DB 255 LKDGDAVAGSMRRLYSHLATASIDPADSGNVCVGNKSGDKVHV-----TYWV 305
QY 544 RIQKRVNTKVLVFNK--KGSV-----FHEKMLRKREPTILEMEFN 585
DB 306 NVLEHASISKGLHDKVSLGATVHFTCDVGNPARNRTWFHNAQPI--HPSSRLTEGN 362
QY 586 IIPLEVGEDPMATELIPSR--SFYRONG-----EPYIGVKASVFLDPRNLS 633
DB 363 VAKITGVWEDSGLYOCVADNGIGFMQSTGRQIIEODSGMKPIYTAAPANIEWMDGFEVT 422
QY 634 TATAAQTDLNFTINDSGTTP--LRTYGMFSVDFREVTSEP--LNAKVKVH----- 681
DB 423 LS-----CNATGVPPVPIHWYGRHL-----ITSHPSQVLRSKPKRSHLFRPDLD 468
QY 682 LDSTQVKMPEHISTYKMLSNLNDPTGLMEEGDPKFE--NQRNKRERTLVNGLREKR 740
DB 469 LEPVYLIMQAGSS--SLSIQAVTLEHAGKYTCENATKHSQSEALV----- 515
QY 741 LFNLDVPESSRCFVKVRAVSERFLPSEOIQ-----GVVSYVNLPRPTGFLSNPRA 792
DB 516 ----VVFETIN-----KAESVTPSEASQNDERPDRGSESSLNLP----- 554
QY 793 WGRFDSVITGPNGACVPARCDOSPDAYS-----AVYLASLAG----- 830
DB 555 ----VKVHPSGVELPAEKNASVDPAPNILSPQTHMDPTNLVVRARCDGMPINAYF 608
QY 831 -----EELQAVESSPK-----FNNMAGVQPIYLNKLIN 858
DB 609 VKYRKLDGSGAVGSMWTVRVPGSENEHLTELEPSLSLEYLVLMVANSVAGEGQPAW--LT 666
QY 859 YRTTHEDPRVAKT--AFQISMAKPRPNASBESNGPIYAFENLRACEEAPPSAHERFYQ 916
DB 667 FRTSKKMASSKNTQASFPVGLPKRPVTAASNS-----NFGV 705
QY 917 IEGDRDYNTVFPNEDDP-----MSWTEDYLAWMPKP-----MEPRACIYKVI----- 960

```

```

DB 706 VLTSSRHSVPEAPDRPTISMASETSVYVTWIPRANGSPITAFKVEYKRMRTSDMLVA 765
QY 961 ---VGP--LEVNVRSSRMGMGTHRTYGKL--YGIRVRSRRDDQDNVSAACLEFPCSG- 1012
DB 766 AEDIPSKLSVEYRSLEPSIYKFRVIAINHG---ESFR-----SSASRPYVAG 814
QY 1013 -----MLYDQDRVPTLV---KVIPGSCSRASVNPMLHEYLVLNPLPAVNN 1056
DB 815 PNFRSMRPTIGRIATEAVSDQIMLKMTYPS-----SNNNTPIQGFYIYRPTDSN 869
QY 1057 DTSEYTMALPDLPLGNHYGIYVTDODPTAKELALGRCPDGTSGSSRIMSNVAVLT 1116
DB 870 D-SDYKR-----DVEEGSKOMHTIGHLQPTSYDIKM-QCFN--EGGESEF--SNVICET 919
QY 1117 FNCVERQVGRS-AFOYLQSTPAQSPAAGTVGGRVSRROQRAS 1159
DB 920 --VKKRVPGASDYPVKELSTPSSSGNAGNVG--PATSPARS 958

RESULT 11
ID 076281 PRELIMINARY; PRT: 6658 AA.
AC 076281; Q24343;
DT 01-NOV-1998 (Tremblrel. 08, Created)
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE BENT PROTEIN (FRAGMENT).
GN BT OR CG1479.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_Taxid=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98300339; PubMed=9636710;
RA Daley J., Southgate R., Ayme-Southgate A.;
RT "Structure of the drosophila projectin protein: isoforms and
RT implication for projectin filament assembly.";
RL J. Mol. Biol. 279:201-210(1998).
RN [2]
RP PARTIAL SEQUENCE FROM N.A.
RX MEDLINE=91376068; PubMed=1910171;
RA Ayme-Southgate A., Vigoreaux J., Benian G., Pardue M.L.;
RT "Drosophila has a twitchin/titin-related gene that appears to encode
RT projectin.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:7973-7977(1991).
RN [3]
RP PARTIAL SEQUENCE FROM N.A.
RX MEDLINE=95146546; PubMed=7844153;
RA Ayme-Southgate A., Southgate R., Saide J., Benian G.M., Pardue M.L.;
RT "Both synchronous and asynchronous muscle isoforms of projectin (the
RT Drosophila bent locus product) contain functional kinase domains.";
RL J. Cell Biol. 128:393-403(1995).
CC -1- SIMILARITY: TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC -1- SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX
CC DOMAIN.
DR EMBL: AF047475; AAC27550.1; -.
DR HSSP: P00518; 1PK.
DR FLYBase; FBgn0005666; bt.
DR InterPro: IPR000255; -.
DR InterPro: IPR000577; -.
DR InterPro: IPR000719; -.
DR InterPro: IPR001245; -.
DR InterPro: IPR001777; -.
DR InterPro: IPR002290; -.
DR InterPro: IPR003006; -.
DR InterPro: IPR003600; -.
DR Pfam: PF00041; f03; 37.
DR Pfam: PF00047; f03; 16.
DR Pfam: PF00069; PKinase; 1.
DR PRINTS: PR00014; FTYPEITI.
DR PRINTS: PR00109; TYRKINASE.

```

DR PROSITE: PS00933; EGGY KINASES.1; UNKNOWN.1.  
 DR PROSITE: PS00012; PHOSPHOPANTHETHEINE; UNKNOWN.1.  
 DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP; UNKNOWN.1.  
 DR PROSITE: PS50011; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE: PS00108; PROTEIN\_KINASE\_ST; 1.  
 DR SMART: SMO0410; IG\_Like; 1.  
 KM ATP-binding; Repeat; Serine/threonine-protein kinase; Transferase;  
 FT Tyrosine-protein kinase.  
 SQ SEQUENCE 6658 AA; 743460 MW; 0B72EB0993F6062E CRC64;

Query Match 2.4%; Score 152; DB 5; Length 6658;  
 Best Local Similarity 19.6%; Pred. No. 0.08;  
 Matches 162; Conservative 100; Mismatches 259; Indels 306; Gaps 40;

QY 234 GGAPASGAII-----YLLKTPKLLTQTD-----SDGRFRIPGLCPDGK---- 272  
 DB 1742 GGASITGYVERKDDPNTGKWKALETSFDPCKARVNDLIAGKQYFRIMAVKAGSKSPS 1801  
 QY 273 --SLIKITKVKFAPIVLMPK---TSIKRATIKAEFVRAETRYMWNPETKARRAGQSVS 327  
 DB 1802 EPSDQMTAKDNFAP-----PKIDRTNIDKITSK-----AGQHTR 1835  
 QY 328 LCCNATGRPRDPKYFMYHNDTL--DPSLY-----KHESKIVLRKIOGHAGEYFCKAQS 380  
 DB 1836 FDIKVSGEF-PATKYWLHNKARLENDSDSYNIDMESYRKILVPTSKRHSKITLKAEN 1894  
 QY 381 DAGAVKSKVAQILIVASDETPCNVPYESTILRLPHDCFQMATNSFYVDGR--CPVKTC 438  
 DB 1895 ESGRDEASFEEVIVLD-----KCPGPEGPLRVY-----DVHKKGCKLKWMA 1934  
 QY 439 GQODNGICRDAVONCCGISTKEEREIOCSGYTLPTKYAKESQ--RCTETSTYGRKVS 497  
 DB 1935 PLDDGGLIDHYI-----IEKMDVE--SGHWLPSGRKESFALNNLEPSHEXKFRYL 1985  
 QY 498 AAD--NCEPMFEGHYVNGNSKVSMTGYKGTFTLHVPODTERLVLTFFYDLRLOKFNVTKY 554  
 DB 1986 AVNTEGESEPLTGEQSVLAKNPFDEPKPGT-----PE-----AVDMQKDHVDLYWR 2032  
 QY 555 LPFNKGSVAVHEIKMLRKKEPT--LEAMETNIIPLGE-----VVGEDPMALLET- 603  
 DB 2033 PPIINDGSGPIIGYV--VEKREKGTDMKIKGTETITPCLGECKATVPLINCEVEFRVVK 2090  
 QY 604 -----PSRSFYRONGEPEYIGV--KASYTFIDP--KNISTATAQDILNF--INDGCD 650  
 DB 2091 AINAGREPSDA-----SKPIITKPKLAPLIDPTKINIRYNEKSGEPIFLDINISGE 2145  
 QY 651 TEPILRT-----YGMFSYDPR 665  
 DB 2146 PAPVTVNQNKKSVQTSFSHIENLPYTKYINNPERKDGKGLKISAHNFYGGDQVVEFO 2205  
 QY 666 DEVISEPLNAG---KKVHLIDSTQVKNP-----EHISTVAKLMSLNDPTGLWEEEDGF 714  
 DB 2206 INIITTKGKPGGPLEVSEVHKDGCKLWKKKPKDGGEFVESYLVKEKPDPIGIMLPVG-- 2263  
 QY 715 KFEQQRNKRREDTFLVGNLEIRERRLFNLDVPSRRCFVAVRYSRRLFPSEIQGV 774  
 DB 2264 -----RSDGPE-----YNDGVLPGHDYK--FR 2284  
 QY 775 ISVINLEPRGTGLSNPRAGRFDSDVITGPNAGACVPACDDSDPAYSAVYLASILEIQ 834  
 DB 2285 VKAVNKE-----GESEPLET-----LGSIIADPF 2309  
 QY 835 AVESSEPFENPAIVGPOPYLKNLKYRRTDHEDPVKKTAFOISNAKPPNSAEESNGPI- 893  
 DB 2310 SVLPKRP---GVPPR-----TDW-----TANKVELAMPBP--ASDGGSPILQ 2344  
 QY 894 -YAFENL-----RACEAPPSAAHFRFYQIEGDRYDNTVTFPNE 931  
 DB 2345 GYIYEVVDKISPLMEKALFETNSPTPTAVOGLLEGNEYQFRVALNK 2391

RESULT 12  
 ID 075851 PRELIMINARY; PRT: 4123 AA.  
 AC 075851;  
 DT 01-NOV-1998 (TREMBLrel. 08, Created)  
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
 DE WUGSC:H\_DJ0751H13.1 PROTEIN (FRAGMENT).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 OC NCBT\_Taxid=9606;  
 RN [1]  
 RA SEQUENCE FROM N.A.  
 RP Leonard S., Graves T., Strommatt C.;  
 RT "The sequence of Homo sapiens PAC clone DJ0751H13";  
 RN [2]  
 RP Submitted (JUN-1998) to the EMBL/Genbank/DBJ databases.  
 RA SEQUENCE FROM N.A.  
 RP Waterston R.H.;  
 RN [3]  
 RA Submitted (JUN-1998) to the EMBL/Genbank/DBJ databases.  
 CC CC  
 CC EMBL, AC004877; AAC36301.1; -  
 CC HSSP; P01130; IAUJ.  
 DR DR  
 DR InterPro: IPR000421;  
 DR InterPro: IPR000561;  
 DR InterPro: IPR000753;  
 DR InterPro: IPR000884;  
 DR InterPro: IPR000923;  
 DR InterPro: IPR001007;  
 DR InterPro: IPR001064;  
 DR InterPro: IPR001092;  
 DR InterPro: IPR001846;  
 DR InterPro: IPR002172;  
 DR InterPro: IPR002223;  
 DR InterPro: IPR002919;  
 DR Pfam: PF000057; ldl\_recept\_a; 11.  
 DR Pfam: PF000090; tsp\_1; 14.  
 DR Pfam: PF00094; vwd; 5.  
 DR Pfam: PF00754; F5\_P8\_type\_C; 1.  
 DR Pfam: PF01826; TIL; 9.  
 DR PRINTS; PR00261; LDLRECEPTOR.  
 DR PROSITE; PS00280; BPTL\_KUNITZ; UNKNOWN.1.  
 DR PROSITE; PS00196; COPERBLUTE; UNKNOWN.1.  
 DR PROSITE; PS00225; CRYSTALLIN\_BETAGAMMA; UNKNOWN.1.  
 DR PROSITE; PS00202; EGF\_1; UNKNOWN.1.  
 DR PROSITE; PS01209; LDLRA\_1; 9.  
 DR PROSITE; PS50068; LDLRA\_2; 20.  
 DR SMART: SMO0011; VWC\_def; 1.  
 KW Glycoprotein.  
 FT NON\_TER  
 SQ SEQUENCE 4123 AA; 434985 MW; 7AAB6FE8DCE012FB CRC64;

Query Match 2.4%; Score 150; DB 4; Length 4123;  
 Best Local Similarity 21.3%; Pred. No. 0.05;  
 Matches 117; Conservative 40; Mismatches 192; Indels 200; Gaps 30;

QY 81 YGDRVCARPLRLLEARTDWTDPAGSTGOVHGSPPREGFW-----C-----LNRE. 123  
 DB 3028 FGAEACGP-TMEAFPSILRC--PGPYPGMCPDCKMLODCAGPASCALISARCTNQT 3084  
 QY 124 QRPQNCNSNTVFLCP---PGSLRDPTEI-----WSPWSEWKSQA 163  
 DB 3085 CHPCCHCPSGMLLTVSRGHPGLGASVQPPVALPGAIGTSVAGAGWMPGMPWHSCSR 3144  
 QY 164 ACGGTGYOTRTRIC-----LAEWVSLCSEASBEGQHMG-----QDCTA----- 202

```

Db 3145 SCG-GLSRRTACDPPPOGLDYCEGPRAOGEVQALPCPTVNTCTAIEGAEYSPGCP 3203
OY 203 -----CD-----LTCPMGOVNMADCA-----CMQODPMLHANVSLPGAPAS 239
Db 3204 CRRSCDDLVHCWRCQPCYCPCGGVLSNGAICVOPGHCCLD-LLTQORHHPGAR--- 3259
OY 240 GAATLTLTKPRLTLQTDSDGRFRIPLGL-CPD---GKSLTKTKVFAFIVLTPMTSLK 295
Db 3260 -----LARPDCNHCCTGEGRLNCTDLPCPDGGGQSL-----HCCGPPCPRS--- 3302
OY 296 AATTAEEVRAETPYVMNPETKARAGOSVSLCCAKT-GKRPDKYFMYHNDTLIDPSL 354
Db 3303 -----CQDLSFGSVCCPG-----SVGCPPTCGCLPG-----LSQDGLCVPPA 3340
OY 355 YKHSKIVLRKIQOHQAGYFPCAKSDAGAVSKYAQLVTASDERPCNPYESYLIRLP 414
Db 3341 H-----CRQYOPGAMAPSFVPSCTVAG-ILQCEVP----- 3371
OY 415 HDGFONATNSFYDVGRCPVKTCAQODNGICRDVAONCGISKTEEREICSGYTLPT 474
Db 3372 -DCPPPGVWSSMGPMEDCSVSGGGEQLSRRC--ARPPCPG-----PA 3412
OY 475 KYAKECSCQCTETRSIVGRV---SAADNGEPMRPFHVYMGNSRYSMTGYGTFT---- 527
Db 3413 ROSRSTSTQVCHREA-GCPAGRLYRCQPGEGCPFSCAHV-----TQVGCSEGECE 3462
OY 528 --LHVPQDT 534
Db 3463 ECHCPEPT 3471

RESULT 13
OY 022631 PRELIMINARY: PRT: 788 AA.
AC 022631.
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DE 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE T21B6.3 PROTEIN.
GN T21B6.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilita; Rhabditoidea;
OC Rhabdilitidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Cottage A.;
RL Submitted (NOV-1995) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., Murray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Koopa A., Saunders D., Shownkeen R.,
RA Smailson N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierly-Mieg J., Thomas K., Vaubin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans."
RL Nature 368:32-38(1994).
DR EMBL; Z68011; CAA92014.1; -.
DR InterPro: IPR000884; -.
DR InterPro: IPR002965; -.
DR InterPro: IPR003609; -.
DR Pfam: PF00090; tsp_1; 6.
DR PRINTS: PR01217; PRICHEXTENS.
DR SMART; SM00473; PAN_AP; 1.
SQ SEQUENCE 788 AA; 88727 MW; BB0D99F1D29BD961 CRC64;

```

```

Query Match 2.3%; Score 147.5; DB 5; Length 788;
Best Local Similarity 27.2%; Pred. No. 0.0069;
Matches 55; Conservative 15; Mismatches 65; Indels 67; Gaps 11;

OY 46 PADTLESPEWMTLTFNIDPVGKGYERLDAIRF-----YYDRVCARPLRL 93
Db 501 PCQTMSEWEMET-----CSASCSGGRER--TRFCLGTNRCEKDYESECSAGCPPEM 554
OY 94 ARTTDTFAGST-GOVYHSGPREGEFCL-----NBRQ--PGONCSNTVRF 138
Db 555 SOMEWEGQCSVTCGQGVAVRQRT---CLGVEFDHLQGPKEQRACDGPSCS----- 604
OY 139 CPGSLRRDTERIWSPSWMSKSAAGOTGYOTRTRICLAEMVSLCSASEGQHCMQ 198
Db 605 -----LWSPWQEMSTCSASCG--SGMKRRQVC--QFGTDCGPNESQFCYGP 649
OY 199 DC-----TACDLTCPMQ 211
Db 650 PCAEWTEWEMSGCSKCGPGQ 671

```

```

RESULT 14
OY 017591 PRELIMINARY: PRT: 1444 AA.
AC 017591.
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-JAN-1999 (TREMBLrel. 09, Last sequence update)
DE 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE C02B4.1 PROTEIN.
GN C02B4.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilita; Rhabditoidea;
OC Rhabdilitidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Mcmurray A.;
RL Submitted (JUL-1995) to the EMBL/Genbank/DBJ databases.
DR EMBL; Z50004; CAA90283.1; -.
DR EMBL; Z50006; CAA90293.1; JOINED.
DR EMBL; Z50006; CAA90302.1; -.
DR EMBL; Z50004; CAA90302.1; JOINED.
DR InterPro: IPR00130; -.
DR InterPro: IPR000884; -.
DR InterPro: IPR001590; -.
DR Pfam: PF00090; tsp_1; 12.
DR Pfam: PF01421; Reprolysin; 1.
DR PROSITE: PS50215; ADAM_MPRO; 1.
DR PROSITE: PS00142; ZINC_PROTEASE; UNKNOWN_1.
DR SMART; SM00209; TSP1; 1.
SQ SEQUENCE 1444 AA; 16067 MW; 32D65E59FABBF75 CRC64;

```

```

Query Match 2.3%; Score 147; DB 5; Length 1444;
Best Local Similarity 28.9%; Pred. No. 0.019;
Matches 44; Conservative 22; Mismatches 64; Indels 22; Gaps 8;

OY 86 CARPLRLERT-----TMTF-AGSTGOVYHSGPREGEFMCINREORPONCSNYTVR-F 137
Db 1297 CFPDPAICHLRDGWSWWSMTWPCASCGFVGQTRDS---CSSPEKGGQSGGLAHQTS 1353
OY 138 LCPGSLRDTERIWSPSWMSKSAAGCQTVQVTRTRICLAEMVSLCSASEGQHCQG 197
Db 1354 LCDLPACDHESDQEWASWMSGCMGCG-IGTRTVACVSPV-----SDGGQPCFG 1406
OY 198 ODTACDLT-CPMGVNAADCAQCMQDFMLHG 228
Db 1407 R---SSEITECROSPSTALCSSFTITSSHLADG 1435

RESULT 15
OY 061541

```

